Høgskulen på Vestlandet

MASTER'S THESIS

Vocabulary knowledge, attitudes towards learning English and language use anxiety: a quantitative study of Norwegian 7th graders

Malin Hopøy

Master's in Education with English Didactics

Department of Language, Literature, Mathematics and Interpreting

Supervised by Monika Bader

01.06.2021

I confirm that the work is self-prepared and that references/source references to all sources used in the work are provided, cf. Regulation relating to academic studies and examinations at the Western Norway University of Applied Sciences (HVL), § 12-1.

ABSTRACT

This research study investigates Norwegian 7th graders' vocabulary knowledge and to the extent to which it correlates with attitudes towards learning English and language use anxiety. Additionally, this study aims to investigate in what way a collaborative output task can affect the pupils' vocabulary knowledge. The study was based on data collected from 74 Norwegian pupils of English in 7th grade over a period of two weeks. A two-group quasi-experimental study was implemented to investigate the effect of a collaborative output task, in the form of a classroom debate. The control group had regular English lessons, while the experimental group participated in a classroom debate project. The pupils were testes twice, where the pre-tests provided information of the pupils' vocabulary knowledge, attitudes towards learning English and language use anxiety. Whereas the post-tests were used to compare results with the pre-test scores in order to investigate a potential effect of the collaborative output task. The tests consisted of a vocabulary level test and an attitude/anxiety survey. The vocabulary level test measured vocabulary frequency levels. Level 1.000 and 2.000 represent the high-frequency words (around 90% coverage of running words), and level 3.000 represent some of the mid-frequency words (Nation, 2013). Considering the coverage of the high-frequency words, it is important that learners of English gain knowledge of at least these levels (Nation, 2013). The value of the test is that it indicates which word frequency learners should focus on.

Results from this study showed that the pupils' vocabulary scores decreased for each level. It also showed that overall the pupils had positive attitudes towards learning English and slightly low levels of language use anxiety. Additionally, the results showed a significant, negative correlation, between both vocabulary knowledge and language use anxiety ($p < .001$) and between language learning attitudes and language use anxiety ($p < .05$). This indicates that the pupils who had low language use anxiety, had high vocabulary knowledge and positive language learning attitudes, and vice versa. No statistically significant correlation was found between language learning attitudes and vocabulary knowledge ($p > .05$).

However, the main finding was the vast variation in pupils' vocabulary scores where some pupils did not master level 1.000, while others mastered level 3.000. This suggests that some pupils will only be able to use survival vocabulary for traveling, while others should be able to read unsimplified text with the help of a dictionary and converse in English. These variations are major; therefore, this thesis stress the importance of teachers assessing their learners' vocabulary knowledge. Once the teachers know their learners' vocabulary knowledge, they can adjust the teaching and the materials to the different frequency levels so that the pupils can keep increasing their vocabulary knowledge through their education.

SAMMENDRAG

Denne forskningsstudien undersøker vokabularkunnskapene til norske syvende-klasseelever og om det kan være en sammenheng mellom dette og holdninger til å lære engelsk og engstelse for å bruke engelsk. I tillegg undersøker denne studien på hvilken måte samarbeids- og produksjonsoppgaver kan påvirke elevenes vokabularkunnskaper. Denne studien har basert seg på data fra 74 norske syvendeklasse-elever over en toukers-periode. En kvasieksperimentell studie, med to grupper, ble brukt for å undersøke effekten av samarbeids- og produksjonsoppgaver, som i denne studien var i form av en klasseromsdebatt. Kontrollgruppen hadde vanlige engelsktimer, mens eksperimentellgruppen deltok i klasseromsdebatt-prosjektet. Elevene ble testet to ganger, hvor førprøven ga informasjon om elevenes vokabularkunnskaper, holdninger til å lære engelsk og engstelse for å bruke engelsk. Etterprøvene ble brukt for å sammenligne resultatene med førprøvene for å undersøke en potensiell effekt av debattprosjektet. Prøvene bestod av en vokabular-nivåprøve og en holdning/engstelse undersøkelse. Vokabular-nivåprøven målte frekvensnivåer. Nivå 1.000 og 2.000 representerer høyfrekvens ord (dekker rundt 90% av løpende tekst) og nivå 3.000 representerer noen av midfrekvens ordene (Nation, 2013). Det er viktig at engelskelever lærer seg høyfrekvensordene med tanke på hvor mye av løpende tekst de dekker (Nation, 2013). Verdien i denne prøven er at den indikerer hvilke frekvensnivå eleven burde fokusere på.

Resultatene fra denne studien viser at elevenes vokabularscore ble lavere for hvert nivå. De viser også at elevene hadde positive holdninger til å lære engelsk. I tillegg viste de lavt nivå av engstelse for å bruke engelsk. Resultatene viser også en signifikant negativ korrelasjon både mellom vokabularkunnskap og engstelse for å bruke engelsk ($p < .001$) og mellom holdninger til å lære engelsk og engstelse for å bruke engelsk ($p < .05$). Dette indikerer at elever som hadde lite engstelse for å bruke engelsk hadde høye vokabularkunnskaper og positive holdninger til å bruke engelsk, og motsatt. Det ble ikke funnet et statistisk signifikat korrelasjon mellom holdninger til å bruke engelsk og vokabularkunnskaper ($p > .05$).

Hovedfunnet i studien var den store forskjellen mellom elevenes vokabularkunnskaper hvor ikke alle elevene mestret nivå 1.000, mens andre mestrer nivå 3.000. Dette indikerer at noen elever kun har overlevelsesordforråd som kan brukes i utlandet, mens andre kan lese autentiske tekster med hjelp av en ordbok og kan konversere på engelsk. Disse forskjellene er store og derfor understreker denne studien hvor viktig det er for lærere å teste elevenes vokabularkunnskaper. Når lærerne vet elevenes vokabularkunnskaper kan de tilpasse undervisningen og materialet til de ulike frekvensnivåene slik at elevene kan fortsette å forbedre vokabularkunnskapene sine i løpet av skoleløpet.

Acknowledgements

When I decided to write a master thesis I had never in my wildest dreams imagined that it would be as challenging as it has been. For a while it felt like I would never finish, but finally I can see the finish line. It truly would not have been possible without all of the wonderful people that helped me along the way.

First and foremost, I would like to dedicate a huge thanks to my brilliant supervisor, Monika Bader. Your clear comments, great suggestions, and belief in me during these last few months made it possible for me to now hand in a thesis that I can be very proud of. That means the world to me. I would also like to thank Matthew Scott Landers for believing in my project and trusting me to handle such an ambitious study. As well as Kristian Andersen Rusten for taking time out of your schedule to give me a lesson in statistical analysis, which was much needed and appreciated.

I would like to thank all of the teachers, parents and principals who allowed me to use their time and pupils/children in this study. You showed great kindness in allowing and trusting me to do this, especially considering that it was in the middle of a pandemic. The pupils who participated in this study deserves a huge thanks. This would obviously not have been possible without you! Not only did you give me valuable data for my study, but you also showed me how much you are able to do, and how willing you are to try out new things. That experience gave me so much joy and it is something I will take with me in my teaching career. Working with you guys was the highlight of this thesis and it made it all worth it.

Lastly, I would like to thank all of the people who have stood by me during my ups and downs of this thesis. It has been a rollercoaster of emotions, and luckily, I have had people around me to comfort me and cheered me on during this process. My lunch-buddies, you know who you are, and you are the best! My mom who has listen to my excitement and frustration of this thesis and my aunt who has given me advice throughout the process. Mohammed Alhassan, Lori Lynn, Maggie Carter and Heather Brooks, who looked at content and proofread this thesis, you are true lifesavers.

This thesis has taught me that it certainly takes a village to write a master thesis. I am beyond blessed to have been surrounded by such an incredible village that made it possible for me to take on this journey and finally reach the finish line. Thank you all from the bottom of my heart!

Table of contents

List of figures

List of tables

1 Introduction

English is often considered a global language, which leads to several implications for learners of English. According to the new curriculum (Utdanningsdirektoratet, 2020), English is central for cultural understanding, communication and identity development. It also states that it is important to know English to be able to take part in the educational, social- and working life. The curriculum, consequently, contains aims for the pupils to achieve competence in English reading, writing and oral communication. To be able to reach these aims the ability and will to use the language needs to be present.

The *ability* to use English has to do with the ability to speak, write, listen and read English. There are several aspects that need to be present in order to use English. Wilkins (1972, p. 111) claimed that "Without grammar very little can be conveyed, without vocabulary nothing can be conveyed". This quote clearly highlights the importance of knowing English vocabulary in order to use the language. In a study by Macaro (2003, p. 62), L2 teachers put vocabulary as the number one topic they would like to research. Meanwhile, Knight (1994, p. 285) points out that students put vocabulary as their number one priority. This indicates that both teachers and learners of English value the knowledge of vocabulary and would like to learn more about it.

The *will* to use the language, on the other hand, has to do with motivation and language anxiety, amongst other aspects. Dörnyei (1998, p. 117) describes the learning process of an L2 as long and often tedious. He claims that "[w]ithout sufficient motivation, even individuals with the most remarkable abilities cannot accomplish long-term goals" (Dörnyei, 1998, p. 117). Motivation provides both the incentive to initiate learning an L2 and the "driving force" to withstand the long process (Dörnyei, 1998, p. 117). Language anxiety is often considered to have a negative impact on language learning. It has been found to cause poor performances on tests and influence both the learning and production of L2 (MacIntyre & Gardner, 1989). Gardner (2010) argues that language anxiety has reciprocal influences on language achievement.

In Norway pupils' learning English as a second (third or foreign) language (L2) need to have the ability and will to keep improving their English vocabulary. The Norwegian curriculum presents aims that the pupils should reach by a certain grade. For 7th grade, some of the aims are: to explore and use speaking patterns in songs and role-play, write coherent texts and express oneself understandably with a varied vocabulary (Utdanningsdirektoratet, 2020). All of the competence aims in the English curriculum demands a certain degree of vocabulary knowledge. Therefore, it is important that the pupils continue to learn new vocabulary through their education. In order to ensure that, the teachers need to measure their pupils' vocabulary

knowledge and adjust the vocabulary teaching accordingly. In that way, the pupils can keep increasing their vocabulary knowledge through their education and gain a sufficient vocabulary to reach the aims for the curriculum.

1.1 Aims and research questions

By establishing the importance of vocabulary in learning an L2, this study was carried out to investigate the vocabulary knowledge of Norwegian 7th graders. Further, the study aimed to investigate motivation and language anxiety as aspects that can potentially impact the vocabulary learning.

Firstly, considering vocabulary, there are several ways to investigate and teach L2 vocabulary. Nation (2013) divides vocabulary up into three levels depending on how frequently the words are used in the language. The high-frequency words are by far the smallest group, and they account for, by far, the largest portion of running words in text. They are estimated to account for around 90% of running words. Considering the frequency, coverage and range these words represent, Nation (2013, p. 13) clearly states the importance of knowing these high-frequency words, as well as to keep increasing the vocabulary knowledge until at least a coverage of the some of the mid-frequency levels is achieved (Nation, 2013, p. 26). Additionally, he presents different ways of dealing with teaching and learning the different frequency levels. Hence, they cannot all be taught in the same way, which suggest that it is important for teachers to investigate their learners' vocabulary level so that the vocabulary teaching can be adapted to the levels the learners are at. In that way, the learners can keep improving their vocabulary throughout their education.

Secondly, motivation and language anxiety are investigated due to their connection to vocabulary learning. Motivation is a huge field within L2 research, and considering the limitations of a master thesis, it was not possible to investigate motivation in its entirety. In Gardner's (2010) research, *attitudes towards learning English, motivational intensity* and *desire to learn English* were elements that together formed the construct *motivation.* In his "Socio-educational Model", motivation has a direct link to language achievements. Hence, motivation should affect the vocabulary learning of the pupils. *Attitudes towards learning English* express the pleasure and enjoyment of learning English, and it was chosen to represent an aspect of motivation for this thesis. From now on, the construct *attitudes towards learning English* will be called L-attitudes. Language anxiety consists of two aspects in Gardner's (2010) work, namely, *language class anxiety* and *language use anxiety*. These two aspects represent

the anxiety associated with the use of English both in and out of the classroom. Together they form the construct *Language anxiety* in this thesis. Language anxiety has, throughout research, been described as the fear of using the language, and in Gardner's (2010) work it has a reciprocal relationship with language achievements. Hence, it can both assist and hinder the L2 learning process, including vocabulary learning. From now on, this construct will be called L-anxiety.

Despite the importance of knowing the learners' vocabulary frequency levels, little attention has been given to frequency levels of elementary grade pupils in a Norwegian context, to my knowledge. The same goes for studies investigating L-attitudes and L-anxiety, there are few studies that provide information regarding these aspects in the Norwegian elementary grade setting and especially their connection to vocabulary learning. Thus, this thesis offers a contribution and insight into the field of receptive vocabulary knowledge, L-attitudes and L-anxiety amongst Norwegian 7th graders, in an attempt to provide information from the Norwegian context that can be useful to the teaching professionals.

Along with the aim to investigate these aspects in the Norwegian 7th grade, there is an additional aspect of this thesis. This thesis also aims to investigate how collaborative output tasks affect the pupils' receptive vocabulary knowledge, L-attitudes and L-anxiety. Collaborative output tasks are activities that encourage the learners to produce output and reflect on and discuss this output collaboratively. Two theoretical perspectives bear directly on collaborative output, namely, Swain's (1985) output hypothesis and Vygotsky's (1978) sociocultural perspective. Swain (1985) argues that the learners need to be provided with opportunities to produce output to successfully learn an L2. Vygotsky (1978) claims that interaction and collaborative work are an essential part of a language learning process. He states that by working collaboratively, the learners will get the opportunity to develop areas beyond what they can master by themselves.

The initial idea to investigate how a collaborative output task affected the pupils came from my own experiences as a pupil. I remember the most engaging and interesting lessons we had were project-based lessons, especially debate, which is an example of a collaborative output activity. However, I did not experience active lessons similar to debate in the English classes until I studied a year abroad. There the psychology lessons were conducted in English and they often included debates. Once I understood how much debate gave me, both in terms of language learning and motivation, I was baffled by the fact that I had never experienced debate in my own English classes in Norway.

This initial motivation and interest in debate lead me to collaborative output tasks and the research that has been done, showing that this is worth investigating. Several studies have investigated the effects of different collaborative output tasks on learners and found that they can be very beneficial for the learners. Previous research suggests a positive relationship between collaborative output tasks and vocabulary learning (Kim, 2008; Nassaji & Tian, 2014; Sun, 2017). Additionally, research has found that speaking in a foreign language, such as collaborative output tasks usually demands, is one of the most anxiety-provoking aspects of L2 learning (Young, 1990), and that collaborative output tasks can be closely related to motivation and language attitudes (Dörnyei & Ushioda, 2011; Henry, Korp, Sundqvist, & Thorsen, 2018)

Although this research provides indications on the effects collaborative output tasks can have on learners, no one in my literature search, has investigated whether a collaborative output task, in the format of a classroom debate, can affect the pupils' vocabulary knowledge, L-attitudes and L-anxiety in the Norwegian context. This thesis, therefore, seeks to contribute to the research on this field from a new angle. Namely, by manipulating the independent variable (debate lessons), which is the variable we believe might *cause* the results, to determine its effect of the dependent variable (vocabulary knowledge, L-attitudes and L-anxiety), which is the variable we measure to investigate whether the independent variable has an effect on it. This was done through conducting a quasi-experimental study over two weeks. The pupils were divided up in to two groups, an experimental group which engaged in the treatment (classroom debate), and a control group without a treatment. Finally, they were tested both before and after the treatment to investigate potential differences.

Overall, this thesis aims to investigate two main aspects in order to contribute to bridging the research gaps in the Norwegian setting. Firstly, by investigating 74 Norwegian 7th graders receptive vocabulary knowledge, L-attitudes and L-anxiety. Secondly, by obtaining information regarding whether a collaborative output task can affect the pupils' receptive vocabulary knowledge, L-attitudes and L-anxiety. These two aspects will be investigated in light of these five research questions:

1. What is the receptive vocabulary knowledge of Norwegian 7th graders?
2. What are the pupils' self-reported language learning attitudes and language use anxiety?
3. What is the correlation between the pupils' receptive vocabulary knowledge, their self-reported language learning attitudes and language use anxiety?
4. Does engaging the pupils in a collaborative output task in the form of a classroom debate have an effect on the pupils' receptive vocabulary knowledge?

5. Does engaging the pupils in a collaborative output task in the form of a classroom debate have an effect on the pupils' language learning attitudes and language use anxiety?

By investigating these five research questions, the main aim is to achieve a deeper understanding of the pupils' vocabulary knowledge and how L-attitudes, L-anxiety and collaborative output tasks can affect the pupils' vocabulary knowledge. This is important to investigate because the pupils' vocabulary knowledge is such a huge part of the pupils' language abilities. Consequently, teachers need to know how to keep improving the pupils' vocabulary knowledge. Additionally, it is important because of the gap in the current research from the Norwegian context.

1.2 Outline of thesis

This thesis is presented through six chapters. The remaining chapters of the thesis are arranged in the following: Chapter 2 provides theoretical background and previous research for this thesis. Chapter 3 describes in detail the methods and materials used in this study, as well as ethical considerations and a discussion of validity and reliability. Chapter 4 presents the results of this study. Chapter 5 discusses the five research questions in light of the findings, theoretical frameworks and previous research, as well as give an overview of limitations of the study and suggestions for further research. Finally, chapter 6 concludes this thesis with drawing together the main findings of this study and offering concluding remarks.

2 Theoretical background and existing literature

This chapter will present theories and previous research. Firstly, a theoretical background for L2 vocabulary, motivation, L-attitudes and L-anxiety will be presented. This will be presented with a main focus on Nation's (2008, 2013) and Schmitt & Schmitt's (2020) work on vocabulary, as well as Gardner's motivational theory (2010, 2019). Secondly, the connection between vocabulary, L-attitudes and L-anxiety in L2 learning will be presented with a focus on Tseng and Schmitt's model of motivated vocabulary learning (2008). Thirdly, an overview of collaborative output tasks will be presented along with its relation to both vocabulary and L-attitudes and L-anxiety through Swain's output hypothesis (1985, 1995, 2005) and Vygotsky's sociocultural perspective (1978). Lastly, a presentation of previous research from the Norwegian context on vocabulary, L-attitudes and L-anxiety.

2.1 Vocabulary

2.1.1 What is a word?

To be able to discuss vocabulary, there needs to be a common understanding of what a word is. A word can be classified and counted in several different ways. Bjørke (2018, p. 179) gives numerous examples of ways to classify words: they can, for example, be long or short, function words or content words, foreign words or native words, etc. These can be further divided where, for example, content words can be divided into nouns, adverbs, etc. We can also divide words according to their frequency in the use of language, which will be the main focus of this thesis and will be discussed more thoroughly in section 2.1.2.

According to Macaro (2003, pp. 63-64) there are four essential word categories that are important to know to be able to count and explore words; *tokens, types, word families* and *lemmas*. *Tokens* refer to the total number of words that are in a text, for example, there are twenty-two tokens in this sentence. *Types* are the number of different words in a text. Measuring types could be useful if we wanted to measure the range of vocabulary used in a text. For example, the sentence "To be able to discuss vocabulary there needs to be a common understanding of what a word is", consists of 18 tokens, but the word *to* occurs three times, and the words *a* and *be* occur twice, hence consisting of only 14 types. *Word families* consist of the headword (e.g., sustain) and all the words that derive from it: sustainable, unsustainable, sustenance etc. Lastly, *Lemmas* consist of a headword (e.g. to run) and its inflected forms: ran, running, runs. It is, therefore, said that the English language has between 400,000-2,000,000 words, depending on how you count them (Schmitt & Schmitt, 2020, p. 8).

2.1.2 What does it mean to know a word?

What does it actually mean to know a word? Is it enough to recognize it in a book, or understand it when you hear it on the radio? Do you need to be able to spell it or explain what it means? Do you have to pronounce it correctly? The vocabulary knowledge of a person can be divided into categories: the breadth and depth of vocabulary and the productive and receptive knowledge of vocabulary.

2.1.2.1 Receptive and productive vocabulary

Receptive vocabulary refers to words we comprehend by receiving language input from others through listening and reading (Nation, 2013, p. 46). Productive vocabulary, on the other side, consists of words that are stored in the mental lexicon which are used when producing speech or writing to convey messages to others (Bjørke, 2018, p. 183).

> Essentially, receptive vocabulary use involves perceiving the form of a word while listening or reading and retrieving its meaning. Productive vocabulary use involves wanting to express a meaning through speaking or writing and retrieving and producing the appropriate spoken or written word form. (Nation, 2013, p. 47)

The receptive vocabulary generally develops before the productive vocabulary, hence the receptive vocabulary is larger than the productive (Schmitt, 2010, p. 21; Schmitt & Schmitt, 2020, p. 37). The terms receptive and productive relate to all aspects of what is involved in knowing a word, as seen in Table 1. This table illustrates and shows what it means to know a word from form, to meaning and use, both regarding the receptive knowledge and the productive knowledge. For example, in the spoken part of English, the receptive knowledge has to do with recognizing a word and know what it sounds like, whereas the productive knowledge has to do with how to pronounce the word.

Table 1
What is Involved in Knowing a Word (Nation, 2013, p. 49)

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

Note: R = receptive knowledge, P = productive knowledge

2.1.2.2 Breadth of vocabulary

The breadth of vocabulary represents how large the vocabulary is and is often referred to as vocabulary size. The breadth is usually measured using word families. It is estimated that the vocabulary size of a native speaker of English should have a range of 10,000-13.000 word families (Schmitt & Schmitt, 2020, p. 11). Luckily, L2 learners do not need to have the same vocabulary size as L1 learners to be proficient users of English (Schmitt & Schmitt, 2020, p. 11). Given the enormous number of English words that exits, it is necessary to prioritize which words to learn depending on the purpose for learning the language. Frequency has proven to be a very useful tool to find the most suitable vocabulary to learn for the specific purpose (Vilkaitė-Lozdienė & Schmitt, 2020, p. 81).

Vocabulary can be divided into three levels according to how frequent the word families are: high-, mid- and low- frequency words. The high-frequency word level consists of the most frequent 1 to 2.000 word families (level 1.000 and 2.000) and they include function and content words (in, for, forest, price). They are the smallest group of words while covering by far the biggest portion of running words in text. High-frequency words with proper nouns etc. are estimated to cover around 90% of text (Table 2). The mid-frequency word level ranges from the most frequent 2.001 to 9.000 word families (level 3.000 to 9.000) and they include generally useful words, which are moderately frequent (angle, anxiety). Mid-frequency words are estimated to cover around 9% of the running words in text (Table 2). The low-frequency word level ranges from 9,001 and up (from level 10,000). They make up the smallest percentage of running words, while being the biggest group of words. These words consist of technical terms for different occupations, hobbies etc. and words that are rarely used in the everyday language.

Low-frequency words are estimated to cover around 1% of the running words in text (Table 2) (Nation, 2013, p. 21). Table 2 shows some of the percentage each frequency level is estimated to cover according to Nation's (2013, p. 21) table.

Table 2

Summary of the Coverage of the British National Corpus by Word Family Lists Made From the Corpus (Nation, 2013, p. 21)

Type of vocabulary	**Level**	**% cumulative coverage of tokens including proper nouns, marginal words, and transparent compounds**
High-frequency	1st 1.000 level	81.14
	2nd 1.000 level	89.24
Mid-frequency	3rd 1.000 level	93.60
	4th 1.000 level	95.37
	9th 1.000 level	98.08
Low-frequency	10th 1.000 level	98.23
	20th 1.000 level	98.86

Nation (2013, p. 24) stresses the importance of learners knowing the high-frequency words and further argues that the time spent on learning them is well justified considering the frequency, coverage and range they represent. The mid-frequency words are also important, but the high-frequency words need to be mastered before moving on to them. Mastery of the Mid-frequency words provide the opportunity to be an independent user of the language (Nation, 2013, p. 18). A seen in Table 3, knowledge of words below level 1.000 provides the possibility to produce the simplest language. However, knowledge of the first 3.000 levels give coverage of basic listening skills, reading and speaking skills. The table further shows that much of the basic skills are covered by the high-frequency levels, and level 3.000, while the higher mid-frequency words provide enough support to be able to take part in more advanced tasks, like reading unsimplified texts without assistance.

Nation (2013, p. 26) argues that learning the 3.000-5.000 word families should be an explicit goal for non-native speakers who know the high-frequent words. Yet, achieving a good coverage of the highly frequent words should be the number one priority. The aims in the Norwegian English curriculum (Utdanningsdirektoratet, 2020) for 7th graders are, for example, being able to listen and understand adapted and authentic texts and express oneself with a varied vocabulary and politeness to the receiver and situation. According to Nation (2013, p. 39) the pupils need to at least have mastery of the high-frequency words to be able to reach aim like

these. It is, therefore, highly important that the teacher investigate the vocabulary level of the learners and focus on getting the learners to reach full mastery of at least the first two-thousand levels. It is, however, important that learners keep increasing their vocabulary knowledge until they at least achieve coverage of the 5.000 level (Nation, 2013, p. 26).

Table 3

Stages Set of Vocabulary-Learning Goals (Nation, 2013, p. 39)

Language use	Number of words	Source of words
Survival vocabulary for foreign travel	120 words and phrases	Nation and Crabbe (1991)
Reading the easiest graded readers	100–400 word families	
Reading intermediate-level graded readers	1,000 word families	
Basic speaking skills	1,200 word families	West (1960: 38–40, 95–134: 'A minimum adequate vocabulary for speech')
Basic listening skills	3,000 word families	
Reading graded readers and using monolingual dictionaries	3,000 word families	*A General Service List of English Words* (West, 1953); BNC/COCA word family lists
Reading mid-frequency readers	4,000/6,000/8,000 word families	BNC/COCA word family lists
Reading unsimplified text with the help of a dictionary, and watching TV	3,000 words	BNC/COCA word family lists
Unassisted reading of unsimplified text	6,000–9,000 words	BNC/COCA word family lists

2.1.2.3 Depth of vocabulary

The depth of vocabulary on the other hand, has to do with how well we know the words. It can also be referred to as the quality of understanding. According to Bjørke (2018, p. 182), the depth has to do with the knowledge of form, content and usage of a word. Exactly when we know a word is not a definable moment, but more on a scale (Schmitt, 2010, p. 16). This scale is usually said to go from receptive vocabulary knowledge to productive, but it is debatable what it takes to reach the productive part of the scale. Read (2000, p. 154) highlights this by asking the question "is there a certain minimum amount of word knowledge that is required before productive use is possible?". Given the difficulties in defining and measuring depth of vocabulary, Schmitt (2014, p. 942) concludes that the construct of depth is too vague to be useful in research. Therefore, for the purpose of this thesis the vocabulary breadth/size in terms of levels will be measured without looking into vocabulary depth.

2.1.3 How are words learned?

Considering the importance of learning vocabulary and the large amount of vocabulary that is needed to do different tasks in English, it is important to understand how to best teach and learn the necessary vocabulary. To understand the different ways of teaching vocabulary this section

will firstly discuss explicit and implicit vocabulary teaching, and secondly show how the different frequency levels can be taught and learned.

According to Macaro (2003, p. 71) the question of how vocabulary is best learned is probably the issue that divides most theorists and teachers. The core of the issue is whether learning new vocabulary should be done explicitly or implicitly. Explicit vocabulary learning implies that there is an explicit focus on words, whereas implicit learning, often called incidental, is when the focus is on the use of language rather than the language itself (Schmitt & Schmitt, 2020).

Schmitt and Schmitt (2020, pp. 140-141) present a summary of several studies which have researched the effectiveness of incidental vocabulary learning through reading. They show different levels of vocabulary pick-up rates from reading, and that vocabulary learning from reading is efficient to develop and enrich partially known vocabulary. They estimate that one needs to be exposed to a word eight to ten times for the learner to attain receptive knowledge of the word (Schmitt & Schmitt, 2020, pp. 143-144). According to Cobb (2007) it is not common to encounter a word beyond the 2.000 most frequent words, eight to ten times in just one authentic text. Consequently, extensive reading provides better opportunities for enough repetition to happen (Schmitt & Schmitt, 2020, pp. 140-142). Extensive reading is a type of persistent, motivated and engaged reading, where the learners read a large quantity for pleasure, information and general understanding (Day & Bamford, 1998, p. 188). Extensive reading has shown positive results when it comes to vocabulary acquisition from incidental vocabulary learning. Pellicer-Sánchez and Schmitt (2010) found that the pupils learned considerable amounts of high-frequency vocabulary from reading an authentic novel. Nevertheless, their findings suggest that the mid-frequency words did not occur enough times for them to be picked up incidentally.

Furthermore, exposure to English outside of the classroom is proven to have clear effects on learners' incidental vocabulary learning. In Iceland, Lefever (2010) found that even before the pupils started with formal English education most of them understood basic spoken English and many could participate in simple conversation. Similar findings were seen in Belgium where a significant proportion of 11-year-olds could already perform tasks at an A2 level (CEFR, (Council of Europe)), before their formal English Education began. This suggests that they should be basic users of English, hence understand simple information and express themselves in known contexts (Council of Europe). They found that the pupils learned English through different media, like video games. This suggests that learning can happen incidentally and especially through activities that interest the pupils. De Wilde, Brysbaert, and Eyckmans

(2020) investigated Dutch secondary school students and found that the three most important types of L2 input for a range of linguistic aspects, including vocabulary, were the use of social media, gaming, and speaking English. They believe that this is due to the social interaction and authentic communication which are offered in these three types of L2 inputs. Sundqvist and Wikström's (2015) study of Swedish pupils shows similar findings, where frequent gamers had the highest vocabulary test scores. Schmitt (2008, pp. 339-340) further supports this by saying that activities that lead to more exposure, attention and manipulation add to the vocabulary learning.

Although several studies find that vocabulary learning can happen incidentally, Schmitt and Schmitt (2020, p. 162) argue that explicit vocabulary learning "almost always leads to greater and faster gains, with a better chance of retention, and of reaching productive levels of mastery". Yet, they conclude that explicit and implicit vocabulary learning positively require each other. They claim that it is impossible to encounter words adequately and to teach all of the contextual types of word knowledge without implicit vocabulary learning. Similarly, words learned through implicit learning may need the added attention from explicit learning to be learned at the productive level of mastery (Schmitt & Schmitt, 2020, pp. 181-182).

According to Nation (2008, 2013), the teaching of vocabulary should be adjusted to the learners frequency-level. As previously mentioned in section 2.1.2, the high-frequency words are very useful for the learners to know since they are such a small group of words (2.000 word families) and still account for such a large portion of running words (90% coverage). It is, therefore, very important that considerable time is spent on learning these words.

According to Nation (2008, pp. 1-3) high-frequency words should be taught and learned through the four strands: *meaning-focused input, meaning-focused output, language-focused learning* and *fluency development*. Meaning-focused input is where learners meet vocabulary through listening and reading and both learn new vocabulary and enrich and establish already met vocabulary. The focus of this strand should be to understand and enjoy the material. Useful activities can be extensive reading, listening to stories and getting involved in conversations. Meaning-focused output involves the learners partaking in speaking and writing activities. The learners should be pushed to use vocabulary at the boundaries of their knowledge through activities like prepared talks, problem solving, conversations and role-play. Language-focused learning consists of the learners deliberately learning vocabulary that can both be new or that they have met before. The teachers should give deliberate attention to vocabulary and vocabulary strategies. Nation (2008, p. 2) mention four specific vocabulary strategies: guessing from context, learning using word cards, using words parts, and using dictionaries. Typical

activities for this strand are intensive reading, learning and practicing strategies and doing vocabulary exercises. Fluency development is where the learners get more proficient at using the vocabulary they already know. During this strand there should only be known vocabulary and the aim is for the learners to practice their reading, listening, speaking and writing skills. "Fluency activities have the characteristics of (1) involving very easy, familiar material, (2) including some pressure to perform faster, (3) having a focus on understanding or producing messages, and (4) often involving repeated use" (Nation, 2008, p. 2).

All of the four strands should get one-quarter of the learning time and the teacher's job is to plan to make sure they get roughly the same amount of time. Furthermore, it is important that there are no large amounts of unknown vocabulary for the meaning-focused input and meaning-focused output strands. If there is, these activities will then become more language-focused than meaning-focused, which reduces the purpose. Similarly, the fluency development activities are not effective if there is much unknown vocabulary (Nation, 2013, p. 3). It is, therefore, very important to know the learners' vocabulary level so that the teachers can adjust both the way to teach, and what material to use, for the different learners. Nation (2013, p. 25) states that the high-frequency words are so important to learn that anything the teachers and learners can do to learn them is worth doing. However, a good way to ensure the learning of high-frequency words is through the four strands which include both implicit and explicit vocabulary learning.

The mid- and low-frequency words on the other hand, should be taught in a different way, since they are usually not encountered often enough in regular text to be learned through meaning-focused input and output. According to Nation (2013, p. 27) the mid-frequency words should be learned through working with vocabulary strategies, where the learners use vocabulary strategies to deal with unfamiliar vocabulary. He suggests four strategies: guessing from context, learning using word cards, using words parts, and dictionary use. Gausland and Haukås (2011) suggest other strategies such as: activities connecting words to rhymes and rhythms, making a mind map and word lists. Teaching the mid-frequency words is not about teaching the specific vocabulary but expanding and refining the learners' ability to learn vocabulary and coping strategies. The low-frequency words should not be given valuable lesson time to be taught. Yet, the focus should be on teaching the learners vocabulary strategies. Learners can also learn this vocabulary incidentally though reading.

Since it is the main goal for learners to know the high-frequency levels, most of the lessons should focus on the four strands and implementing them. Once the learners have enough mastery of the high-frequency words the teacher can move on to teaching strategies for the mid-

frequency words. Table 4 illustrates how the learning process within the different frequency levels should be structured, where there is a change from the high-frequency words to the mid-frequency words on who is responsible for giving attention to words, while both the teachers and learners should give attention to vocabulary strategies for both frequency levels.

Table 4

The Differing Focuses of Teachers' and Learners' Attention to High- and Mid-frequency Words (Nation, 2013, p.28)

	High-frequency words	Mid-frequency words
Attention to each word	Teacher and learners	Learners
Attention to strategies	Teacher and learners	Teacher and learners

2.1.4 Measuring vocabulary

There are several ways to measure vocabulary. For this thesis it was chosen to measure receptive vocabulary knowledge. According to Schmitt (2010, p. 38), very little research goes beyond the receptive vocabulary. This is not very surprising, seeing as it takes a lot more time and effort to measure productive vocabulary. A productive vocabulary test would have to include free language production either orally or in writing. The time it takes to administer and assess such tests makes them difficult to implement. Receptive vocabulary tests on the other hand, offer researchers more control and they are easy to implement and assess. Since receptive vocabulary tests are easier to implement, it also means that they are easier for teachers to use during their lessons to assess their learners' vocabulary knowledge. Hence, given the limited time of this thesis, the clearer boundaries of receptive vocabulary test and the potential interest for teachers, measuring receptive vocabulary seemed like the best fit for this thesis.

When it comes to the breadth and depth of vocabulary, it was decided to measure the pupils' vocabulary breadth based on the discussion in section 2.1.2. Measuring vocabulary breadth can be done by measuring the total amount of known words (vocabulary size), or the frequency levels of known words (vocabulary level). As explained in section 2.1.2, it is very important for learners to know the high-frequency words, and as seen in section 2.1.3, the different frequency levels should be taught differently, hence making it highly valuable for the teachers to know the learners' vocabulary level to be able to adapt the teaching respectively. The vocabulary level tests are quick to administer and easy to interpret (Nation, 2013, p. 36). Thus, being a good test for teachers to use to investigate what level the learners need to be working on, and roughly how much work needs to be done on these frequency bands before moving on to the next level.

"The Updated Vocabulary Levels Test" developed by Webb, Sasao and Ballance (2017) was chosen as the vocabulary test for this thesis. Its main purpose for this thesis was to pick up the pupils' vocabulary level both before and after the treatment. Further details of the test can be found in section 3.5.

2.2 L2 attitudes, anxiety and motivation

Motivation is seen as a key factor that influences the rate and success of L2 learning in general (Dörnyei, 1998, p. 117). Motivation specifically has an important role in the vocabulary learning process, where motivation should, for example, be present to expect the learners to notice the words in the first place (Nation, 2013, p. 102). Dörnyei (1998, p. 117) claims that motivation provides the incentive to initiate the L2 learning, as well as the driving force to sustain the "long and often tedious learning process". He further argues that even the most outstanding learners will not be able to accomplish long-term goals without sufficient motivation. 22 years ago, Scheidecker and Freeman (cited in Dörnyei, 2001, p. 1) stated that "motivation is, without question, the most complex and challenging issue facing teachers today". The number of articles published on L2 motivation during the last decades suggest that this is still true. Motivation is something that affects all of us, all the time, hence, making it a quite complex and complicated area to investigate (Dörnyei, 2020, p. 61). Dörnyei and Ushioda (2011, p. 3) claim that motivation is about "[w]hat moves a person to make certain choices, to engage in action, to expend effort and persist in action", this might seem straightforward, but the research and debates regarding this provide few straightforward answers.

The enormous field of motivation in L2 makes it difficult to explore in its entirety alongside vocabulary in such a small-scale study, as this thesis. Therefore, two areas within the field of motivation were investigated in this thesis to represent the bigger picture: namely, attitudes towards learning English (L-attitudes) and language anxiety (L-anxiety). These two areas were chosen because of the link they have to vocabulary learning, in addition to their connection to the treatment, debate, which will be exemplified through the following sections.

2.2.1 Gardner's motivation theory

The foundation of L2 motivation research comes from the work of social psychologists working in Canada in the late 1950´s with Robert Gardner and Wallace Lambert at the forefront. Their work characterized the period from 1959 till 1990, called, the social psychological period. During this period, Gardner and Smythe (cited in Gardner, 2010, p. 81) put forward the original

version of “the socio-educational model of second language acquisition”. In this model it is proposed that the individual’s willingness and ability to take on features of another community is an important part of the L2 learning process. In the beginning the learners might just be counting in another language, but with time, the learners might start to think and act as members of that language community. This makes the learning process of L2 unlike any other school subjects (Gardner, 2010, pp. 2-3). A simplified version of ‘the socio-educational model of second language acquisition’ is illustrated in Figure 1.

Additionally, Gardner made a tool, based on the socio-educational model, to test for motivation in the classroom, called, Attitude/Motivation Test Battery (AMTB; Gardner, 1985). This is the measuring instrument that the current study draws on (overview in Table 5). The original AMTB is a motivation questionnaire made up of over 130 items. It was originally made to investigate the acquisition of French as an L2 for English speaking Canadians (Gardner, 2019, p. 23). In the early 2000s, a version was made for English as a foreign language, which was later translated and used in other countries like Japan, Poland and Spain (Gardner, 2019, p. 24).

Although the test battery and model are not without issues, which have been discussed in several articles and later clarified by Gardner (2010), it has set high research standards for the field. It was by far the most commonly employed motivation battery in the 1980s and ‘90s (Dörnyei & Ryan, 2015, p. 77). Figure 1 and Table 5 show how the socio-educational model and the AMTB are related, where the categories of the AMTB are drawn from the model.

Figure 1

A Structural Equation Representation of the Socio-educational Model from Gardner (2006, cited in Gardner, 2010, p. 88)

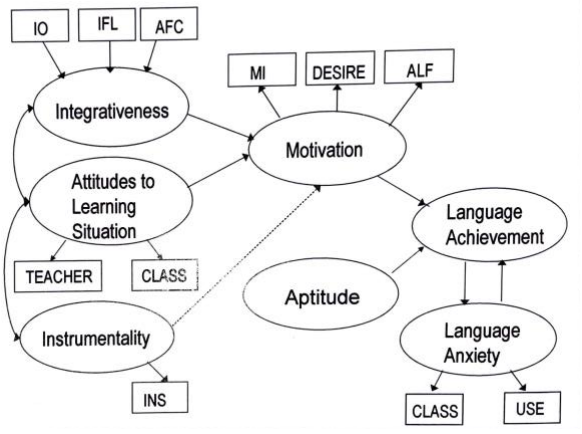


Table 5
The AMTB Developed for Learning English as a Foreign Language (Gardner, 2010, pp. 114-129)

Construct	Measures	Number of items	Example of item
Integrativeness	Integrative orientation	4	Studying English is important because it will allow me to be more at ease with people who speak English
	Attitudes toward English speaking people	8	I wish I could have many English speaking friends
	Interest in foreign languages	10	I wish I could speak many foreign languages perfectly
Attitudes toward the Learning Situation	English teacher evaluation	10	The less I see of my English teacher, the better
	English course evaluation	10	I look forward to the time I spend in English class
Motivation	Motivational intensity	10	I really work hard to learn English
	Desire to learn English	10	I sometimes daydream about dropping English
	Attitudes toward learning English	10	I really enjoy learning English
Language Anxiety	Language class anxiety	10	It embarrasses me to volunteer answers in our English class
	Language use anxiety	10	It doesn't bother me at all to speak English
Instrumentality	Instrumental orientation	4	Studying English is important because it will be useful in getting a good job
Other attributes	Parental encouragement	8	My parents try to help me to learn English

Figure 1 shows how the different aspects of the model are related to each other. For this thesis, the focus will be on the connections between motivation, language anxiety and language achievement, which all have direct links to each other in the model. In the model, vocabulary is a part of language achievement. The motivation construct comprises three elements, motivational intensity (MI), desire to learn English (DESIRE) and attitudes toward learning French (ALF). Attitudes towards learning English (French in the model) was chosen as the component in this study to represent motivation. It is not considered sufficient by itself to measure the entirety of motivation, but it can express the pleasure and enjoyment associated with the process (Gardner, 2010).

L-anxiety on the other hand, is not directly linked to motivation in the model, but serves as a facet of motivation. The model indicates that language achievements and language anxiety have reciprocal influences on each other. This reciprocal relationship makes it a very interesting

aspect to study alongside vocabulary knowledge in this thesis. L-anxiety in the model consists of two aspects, the anxiety one has in the classroom, and the anxiety one has when using the language outside of the classroom. L-anxiety is independent of general anxiety, where L-anxiety develops in individuals as a response to learning and attempting to use the language (Gardner, 2010, p. 91). MacIntyre, Ross and Clément (2019) categorize L-anxiety as an emotion, and according to Izard (2007, p. 273), emotions can be viewed as the primary motivational system for human behavior. Hence, linking L-anxiety and motivation together on a closer level than the socio-educational model does. L-anxiety can come across as fears, and especially for situations where one produces language: fear of speaking, fear of being misunderstood, fear of being laughed at etc. (Dörnyei & Ryan, 2015, p. 176). From the AMTB, both English class anxiety, and English use anxiety were part of the survey submitted to the pupils in this thesis.

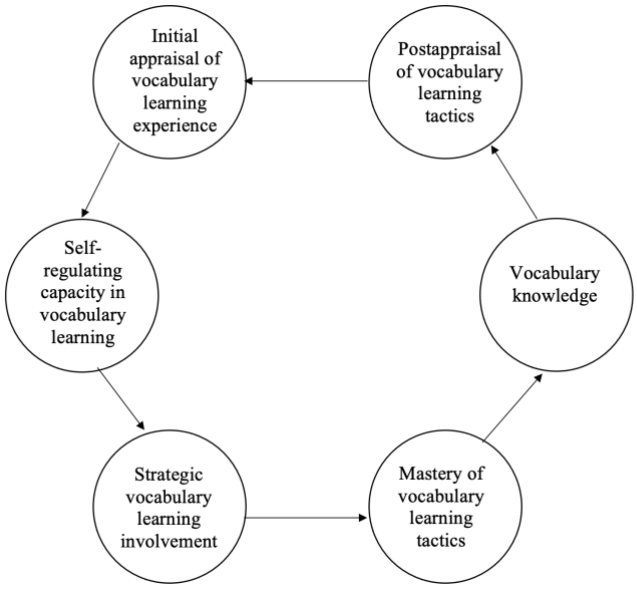
2.3 Vocabulary and attitudes/anxiety in L2 learning

As shown in Figure 1, there are direct links between vocabulary learning (language achievements) and L-anxiety/L-attitudes. Regardless, this has not received much attention in L2 research (Hulstijn & Laufer, 2001; Tseng & Schmitt, 2008).

A study by Tseng and Schmitt (2008) has, however, tried to fill this gap. They conducted a study to provide a heuristic model of the relationship between vocabulary and motivation, where motivation consists of language attitudes, anxiety and self-efficacy. They state that motivation is multidimensional and rarely remains constant in practice; it goes "[…] through a number of interconnected processes in terms of initiating, maintaining and reflecting upon acts of learning in a task" (Tseng & Schmitt, 2008, p. 359). They further claim that this dynamic perspective of motivation is especially relevant for vocabulary learning, given that attaining an adequate vocabulary depth and breadth is such a long and tedious learning process. To investigate the link between motivation and vocabulary, Tseng and Schmitt created and tested a "model of motivated vocabulary learning", which is provided in a simplified version in Figure 2. They found that the vocabulary-learning process is cyclical in nature, where it is essential to go through each stage to move on to the next. Hence, implying that motivation is not just an initial state factor, but an integral part of all stages (instigating, sustaining, and evaluating) that drives the vocabulary cycle along.

Figure 2

A Simplified Version of the Model of Motivated Vocabulary Learning (inspired by Tseng & Schmitt, 2008, p. 381)



The model starts with the variable "initial appraisal of vocabulary learning experience" which is conceptualized as the initial motivational level of vocabulary learning. The measures of this variable are made up of three indicators: vocabulary learning anxiety, vocabulary learning attitude, and vocabulary learning self-efficacy. Tseng and Schmitt (2008, p. 370) explain that they chose learning anxiety and self-efficacy as the indicators because they represent the learners' overall self-confidence, hence being able to reflect the learners' perceptions of achievement. Further, learning attitudes were chosen because they capture the learners' perception of vocabulary learning and "represent a theoretical integration between psychology-based (i.e., Ajzen's Theory of Planned Behavior; Ajzen, 1988) and language-based (i.e., Gardner's social-psychological approach; Gardner, 1985) motivational theory" (Tseng & Schmitt, 2008, p. 370). This variable, initial appraisal of vocabulary learning experience, affects the learners' self-regulating capacity, which is an aptitude that can be developed and influenced by experiences. Tseng and Schmitt (2008, p. 384) claim that when the learners master a set of vocabulary learning tactics, they both have the skills and will that is necessary to achieve their learning goals. The *will* comes from motivation which derives from the initial appraisal of the vocabulary learning experience. They further argue that the vocabulary learning performance also influences the initial motivational state of vocabulary learning, hence, making it a cyclic construct.

This model shows that motivation and vocabulary both influence each other throughout the entire process. It suggests that it is necessary for the learners to have motivation throughout the process to withstand the tedious learning of vocabulary. The model also highlights the integral part language attitude and anxiety have in the cycle.

MacIntyre and Gardner (1989) conducted a study that can further shed some light on the connection between vocabulary and L-anxiety. Their study was conducted on 104 English native speakers from the age of 18 to 25. The participants completed several anxiety scales, a French-English paired-associates learning task and a French vocabulary production test. The results from the scales and tests indicated that L2 anxiety can influence both the learning and production of vocabulary and that L2 anxiety can cause poor performance on the vocabulary tests. More recently, MacIntyre (2017, p. 17) presented a table summarizing the different effects of L-anxiety, where the following were mentioned: length of time studying new vocabulary items, memory of new vocabulary items, time required to complete a test of vocabulary and retrieval of vocabulary from long-term memory. These findings further illuminate this issue of the connection and impact anxiety can have on vocabulary learning, retention and production.

2.4 Collaborative output tasks

The treatment implemented in this study, a classroom debate, consists of three components: reading, collaboration, and producing output. A classroom debate can be viewed as an instance of what in the literature is referred to as, collaborative output tasks. As mentioned in section 2.1.3, studies have shown that reading can be beneficial when it comes to vocabulary learning. Therefore, this section will present connections between collaborative output tasks, and vocabulary learning, L-attitudes and anxiety. This will be presented through two theoretical perspectives that will provide the theoretical rationale for collaborative output tasks. Firstly, Swain's output hypothesis, which argues that learners need to engage in language production to increase their L2 proficiency, and secondly, the sociocultural perspective, based on Vygotsky's ideas of how learning takes place in the mind through social and collaborative interactions.

2.4.1 Output hypothesis – Swain (1985)

Input and output are both essential for L2 acquisition. The specific roles they have in language learning is, however, debated amongst researchers. Krashen (1981, p. 107) argues that one can learn an L2 from input only, and that you can acquire an L2 without ever having to produce it.

Swain (1985), on the other hand, argues that learners need to produce output, in addition to input, to successfully acquire an L2. Swain's output hypothesis outlines three functions of output in L2 acquisition: 1) a noticing (or triggering) function; 2) a hypothesis testing function; and 3) a metalinguistic function (Swain, 2005, p. 471).

The noticing function suggests that when L2 learners engage in producing output (writing or speaking), they become aware that they do not know how to say (or write) what they want to say. Hence, noticing a gap in their linguistic ability. This may trigger certain cognitive processes that have implications for L2 learning, for example, searching for new information to fill this gap (Swain, 2005, p. 474). Several studies have investigated the noticing function of output and provided empirical evidence of its existence and relationship with L2 learning (Izumi, 2002; Swain & Lapkin, 1995).

The hypothesis testing function is based on how output can function as a "trial run", reflecting the learners' hypothesis of how to say or write their meaning in L2 (Swain, 2005, p. 476). This function provides opportunities for the learners to figure out the meaning of words, and how to say them, by trying it out. The learners will modify their output both from trying to express themselves, and the feedback they receive while trying. When a learner says something incorrectly, it is often an indication that they have formulated a hypothesis about how the language works, and are now testing it out (Swain, 1995, pp. 130-131).

The last function, the metalinguistic function, posits that "opportunities for output encourage learners to consciously reflect upon language, thinking about what to say and how to say it" (Nassaji & Fotos, 2011, p. 105). Swain (1995, p. 132) explains that under certain task conditions, the learner will not only reveal their hypotheses, but reflect on it, with language. This level of output represents the metalinguistic function of using language to reflect on language, which allows the learner to control and internalize it.

To summarize, according to the output hypothesis of L2, output tasks are highly valuable activities. When the learners produce language, they may notice a gap in their linguistic abilities, test their hypothesis about the language, and use language to reflect on language, which is necessary to successfully acquire an L2.

2.4.2 Sociocultural perspective - Vygotsky

In a classroom debate, such as the one implemented in this study, group work is a major part of the activity. Vygotsky's sociocultural theory highlights the importance of interaction and collaborative work in the process of language learning (Nassaji & Fotos, 2011, p. 107). When learners collaborate with others, they can develop areas which they have not yet mastered

independently. A concept that describes this process is the zone of proximal development (ZPD), which refers to the distance between the learners' independent developmental level and the level of potential developmental determined through interaction with more capable peers or adults (Vygotsky, 1978, p. 86).

Nassaji and Fotos (2011) point out two other key concepts that are central to the Vygotskian sociocultural theory: *scaffolding* and *regulation*. Scaffolding refers to the supportive environment that is created through the guidance and feedback learners receive during collaboration within the learners ZPD (Donato, 1994). Hence, when learners collaborate with more capable peers or adults, they have the opportunity to master what they have not been able to master on their own (Nassaji & Fotos, 2011, p. 106). Scaffolding happens within the ZPD to move the learner closer to the next stage of development. The notion of regulation explains how "new knowledge begins in interaction and becomes internalized and consolidated through interaction and collaboration" (Nassaji & Fotos, 2011, p. 107). To summarize, collaborative tasks are valuable, according to the sociocultural theory, because when learners collaborate, they have the opportunity to further develop what they have already learned independently (Nassaji & Fotos, 2011, p. 107).

2.4.3 Collaborative output tasks and vocabulary learning

The sociocultural theory and output hypothesis provide arguments which suggest that working collaboratively to produce output is beneficial for learners in developing their L2 knowledge. Vocabulary is a part of L2 acquisition in general, but it has not received much specific attention in research on collaborative output tasks (Kim, 2008, p. 118). Nevertheless, some studies report positive relationships between collaborative output tasks and L2 vocabulary learning (Kim, 2008; Nassaji & Tian, 2014; Sun, 2017). Nassaji and Tian (2014) found that learners who worked collaboratively with output-based tasks produced significantly more correct English target words than the learners who worked independently with input-based tasks. Sun (2017, p. 108) found similar results, where several different instructional modes were investigated and showed that reading plus, either vocabulary instructions or collaborative output, was more helpful for vocabulary acquisition than reading alone. Additionally, reading plus collaborative output was superior when it came to vocabulary retention. These studies show that there is a relationship between collaborative output activities and vocabulary learning. Given the scarcity of research on this topic, it seems that this relationship is worth exploring further.

2.4.4 Collaborative output tasks and language anxiety and attitudes

Collaborative output tasks include a production of language, often orally, while working with others. However, speaking in a foreign language can be very anxiety-provoking, and according to Young (1990, p. 539), learners report that as being the most anxiety-producing part of L2. Horwitz, Horwitz and Cope (1986, p. 127) identified three components that were conceptually related to L2 anxiety which were: communication apprehension, fear of negative evaluation and test anxiety. Communication apprehension is the fear or anxiety about communicating with people, hence, being closely linked to collaborative tasks. This can come across as difficulty to speak in groups (oral communication anxiety), in public ("stage fright"), or in listening to or learning a spoken message (receiver anxiety).

At the same time, several studies have provided information about how the speaking process can be less anxiety-provoking for learner. For example, by being able to prepare and practice in advance (Byrne, Flood, & Shanahan, 2012, p. 576; Ozturk & Gurbuz, 2014, p. 14; Young, 1990, p. 545), work in groups (Young, 1990, p. 543) or being in a warm and friendly environment (Hashemi, 2011, p. 1813; Young, 1990, p. 550).

Consequently, oral production tasks can be the most anxiety-provoking activities for learners, and at the same time studies have found that including a collaborative aspect can reduce anxiety. This has great implications for this thesis, seeing as the pupils will be engaged in a collaborative output task where they will produce oral output. On the one side, the pupils are set up to take part in a highly anxiety-provoking activity, but on the other side, they get to work in groups and prepare in advance. Hence, making it very interesting to see if the pupils' L-anxiety will be affected by participating in this task over such a short period of time.

When it comes to the effect collaborative output tasks have on attitudes towards learning English it seems relevant to look at how working collaboratively can affect language motivation in general. Although not a lot of studies have investigated the link between collaborative output tasks and L2 motivation, Dörnyei and Ushioda (2011, p. 28) mention a connection between collaborative work and motivation. They claim that when individuals work together in groups, their motivational levels can significantly exceed the levels they would have demonstrated if they had worked independently. Dörnyei (1994, p. 278) also states that group-specific motivational components is one out of three motivational components that are specific to learning situations. Henry, Korp, Sundqvist and Thorsen's (2018, p. 267) research on Swedish L2 teachers found that many teachers mentioned projects where students collaboratively produced output over a longer time period to be a highly motivating activity. They further state that when projects work, they get entire classes caught up in a tide of motivational energy,

where enthusiasm and goal-targeted behavior becomes the focus (Dörnyei et al, cited in Henry et al., 2018, p. 268). These studies suggest that working collaboratively to produce output can greatly affect the learners' L2 motivation, hence affect the learners' attitudes towards learning English.

2.5 Research on vocabulary, L-attitudes and L-anxiety in the Norwegian context

Considering the vast amount of research that has been conducted both on vocabulary and motivational aspects in the English L2 classroom, it was considered necessary to limit the scope when presenting previous research. This section will, therefore, present research conducted in Norway. This was also decided in order to find studies which are conducted under similar language circumstances as this study. Although vocabulary and motivation are widely studied areas on the international level, less research has been conducted in the Norwegian context. Several of the studies that are found have commented on this research gap, and their studies, along with this, lay the foundation to limit this gap.

The first study is by Langeland (2012). She looked at how the English vocabulary of forty 5th graders developed over three years. She tested the pupils receptive and productive written vocabulary and found that the pupils moved slowly, but not evenly, towards greater lexical richness both in their receptive and productive vocabulary through the three years. The receptive vocabulary test (Peabody Picture Vocabulary Test Version 4) showed an increase in average scores from 122.3 to 149.4 from 5th till 7th grade. The productive writing tests showed that while the pupils had a decrease in number of tokens written from grade 5 to 6, the number of types and families almost doubled. In 7th grade tokens, types and families trebled from 5th grade. Similarly, the pupils started to use more lower-frequency words the higher the grade. Langeland does, however, point out that some pupils had very low scores and a slow progress, hence urging teachers to test pupils' vocabulary yearly to identify those children.

Helness (2012) conducted a study looking at 57 written texts produced by 7th graders and 57 written texts by 10th graders and compared their vocabulary variation, lexical density and text length. She did not find a difference in lexical density between the two groups, and the vocabulary variation seemed to be higher in grade 10, but these results were found to be inconclusive. Yet, 7th graders wrote longer texts, despite having less time than 10th graders, which Helness implies can be due to the nature of the writing tasks. The 7th grade task was targeted more closely to their level, whereas the 10th grade task was targeted at higher levels.

Two master's theses help shed light on the vocabulary knowledge of Norwegian pupils in lower and upper secondary schools. Olsen's (2016) master thesis focused on how an explicit focus on vocabulary can facilitate pupils' vocabulary acquisition. He investigated this through a quasi-experimental study conducted with Norwegian lower secondary pupils, where the experimental group received teacher-initiated focus on vocabulary and the control group received additional focus on content, while working with texts about global warming and climate change. Olsen found that the experimental group acquired more vocabulary items compared to the control group from a targeted keywords test. Additionally, he tested the pupils' overall vocabulary size at the beginning of the study with the Vocabulary Size Test by Nation. He found that out of the forty-one participants, the control group had a vocabulary size with an average of 7998, while the experimental group had an average of 5400.

Onyszko's (2019) master thesis investigated the receptive and productive vocabulary size of 100 students from an upper secondary school in Norway. Through the Vocabulary Size Test by Nation, she found that the VG1 students had an average receptive vocabulary size of 8,338 word families and an average of 4,769 word families for their productive vocabulary size. Additionally, she found that the students heavily relied on high-frequency words when writing in English and that the mid- and low-frequency words were used very limited.

These studies on vocabulary focus mostly on measuring the pupils' vocabulary size and not size according to level, which this study does. Even though some of them mention levels and frequency, none of the studies have used a vocabulary levels test. Hence, presenting a gap in the research conducted in Norway, especially considering the great importance of knowing the learners' vocabulary level.

Motivation is also a widely researched area within L2 learning. But similarly, there is not a lot of research from the Norwegian context. Some have, however, tried to fill this research gap with their contributions to the field. Jakobsson (2018) found as part of his study that Norwegian 10th graders motivation towards English correlated with their grades in written and oral English. The pupils who reported having negative attitudes towards English received low English grades, and the pupils who reported positive attitudes attained high English grades.

Gjerde (2020) wrote a master thesis on the issue of L-anxiety in oral activities in the English lessons in Norway. She investigated lower secondary English teachers' and pupils' beliefs regarding this. She found through interviews and questionnaires that both the teachers and pupils considered high self-exposing activities that included oral communication to be the most anxiety-provoking activates in the English classroom. Gjerde explains that the pupils reported the fear of making mistakes, being critically evaluated and mocked, as the main

reasons for their classmates' L-anxiety. However, some aspects helped reduce the L-anxiety, such as lower self-exposing activities, like group work, speed-dating and group games. Additionally, the pupils found it anxiety-reducing when the teacher primarily spoke English during the lessons and if the teacher was calm and patient.

Lastly, a study conducted by Myhre and Fiskum (2020) where twenty-three Norwegian 8th graders were interviewed about their development of spoken fluency of English in an outdoor environment. They found that the pupils' L-anxiety decreased when they had lessons outside in smaller groups. The pupils reported that making mistakes in oral English was one of the main reasons for why they were anxious, hence working in smaller groups reduced the L-anxiety since not as many people heard you making the mistakes. The outdoors also helped reduce L-anxiety because of the lesser theoretical focus they experienced outside, where they could speak more freely without worrying about the accuracy of what they said. Myhre and Fiskum conclude that the pupils reported an increased willingness to communicate in English when they could use the language in a more realistic setting, and when they had interesting ways of learning.

None of the studies focusing on L-anxiety have used tests or surveys on a larger group of learners to investigate their L-anxiety, but rather interviews on smaller groups. Hence, resulting in a research gap of larger, more generalizable data collected regarding the pupils' L-anxiety. Further, there is a gap in the research conducted in Norway on the relation between vocabulary and L-attitudes and L-anxiety, where no explicit studies were found. Given the limited relevant research found from the Norwegian context, this study can help shed a light on these areas within the L2 research, and hopefully contribute to the filling of this research gap.

3 Methodology

The main aim for this study was to investigate the vocabulary knowledge of Norwegian 7th graders and to investigate motivation and language anxiety as aspects that can potentially impact the vocabulary learning. This chapter will present the methodology used to investigate these aims. The research design and procedures of the study will be presented, along with comments on the participants and the material and measuring instruments. Further, a short explanation of the statistical methods is presented. Lastly, ethical considerations and threats to the validity and reliability of the study will be discussed.

3.1 A quantitative research design

This study aims to obtain general results that can be useful for teachers and future researchers beyond the participants of this study. In order to reach these aims through the research questions, the results need to be outcome-oriented, meaning that the interest lies in the scores the pupils obtain, not how they feel or think about achieving those scores. Firstly, getting general results about Norwegian 7th graders receptive vocabulary knowledge, L-attitudes and L-anxiety, and secondly about how these aspects are affected by a classroom debate treatment. In order to reach these aims and get results that can be generalized for the larger population, and not just for the selected sample, the results have to be an accurate and reliable description of reality.

According to Dörnyei (2007, p. 31), a quantitative research approach offers the possibility to explore questions in an objective manner to accurately present a description of the world through the use of numbers. Mackey and Gass (2015, p. 4) provide an overview of characteristics dividing quantitative and qualitative research, where main differences are, for example, objective vs. subjective, outcome-oriented vs. process-oriented, hard and replicable data vs. soft data and generalizable vs. ungeneralizable. Hence, the use of a quantitative research design was implemented, to be able to achieve the aim of providing information that can be generalizable for teachers across the nation.

Further, this study goes beyond the collection of test results, it also implements a treatment and two separate group conditions. This type of a quantitative research is called an experimental design, which can determine unambiguous cause-effect relationships (Dörnyei, 2007, p. 115). A true experimental design is done when two (or more) groups are investigated under tightly controlled conditions and the two groups receive different treatments. The aim is to investigate a cause-effect relationship. The progress of the two groups is usually measured by implementing pre-tests (before the treatment) and post-tests (after the treatment). The only

difference between the two groups needs to be the treatment. In that way, if a difference occurs between the two groups, it is due to the treatment. The best way to do this is to have randomized groups, which should make the average participant in one group comparable to the average participant in the other group. Two comparable groups will have different scores due to a difference in treatment (Dörnyei, 2007, p. 116). This is, however, rarely possible to achieve in most educational settings because of practical constraints. Consequently, a quasi-experimental design was implemented for this study to look for a cause-effect relationship without having randomized groups.

A quasi-experimental design is similar to the classical experimental design with the exception that it lacks the random assignment. When pupils are randomly assigned to groups, they have an equal opportunity of being selected as any other individual. Such random assignment strengthens the validity of the results because it limits the threats of, for example, initial group-differences. Nevertheless, quasi-experimental studies have been accepted where randomization is impossible or impractical, such as it is in this study, where there was only one researcher with limited time and resources (Dörnyei, 2007, p. 117). This study will, therefore, use a quasi-experimental design to investigate the potential effects of a classroom treatment on the vocabulary knowledge, L-attitudes and L-anxiety.

Table 6 presents an overview of the study. It is visible that the only difference between the groups is the treatment. The pre-test from both groups forms the basis for the first three research questions, regarding the receptive vocabulary knowledge, L-attitudes and L-anxiety of Norwegian 7th graders. A comparison between the scores of the two groups and a comparison between the individual groups pre- and post-test scores are used to answer the last two research questions, regarding how a classroom debate can affect the vocabulary knowledge, L-attitudes and L-anxiety of Norwegian 7th graders.

Table 6
Overview of Study

Group	Week 1	Week 2	Week 3	Week 4
Control	Pre-test	Regular English lessons x3	Regular English lessons x3	Post-test
Experimental	Pre-test	Classroom debate lessons x3	Classroom debate lessons x3	Post-test

3.2 Pilot study

It was necessary to run a pilot study, considering that no previous studies were found using the same tests or design in a Norwegian context. According to Mackey and Gass (2015, p. 52) “A

pilot study is an important, if not essential, means of assessing the feasibility and usefulness of the data collection methods and making any necessary revisions before they are used with the research participants". A pilot study was, thus, conducted to test the design and measuring instruments of the study.

The two main aims of the pilot study were first to check that the topic of the debate was successful with 7th graders and second to see how the tests worked with the pupils. It was decided that both aims could be achieved with a smaller-scale pilot study. The pilot study was, therefore, conducted in one 7th grade classroom over one week. This class was not part of the rest of the study, it was a class from the western part of Norway, and it only functioned as a pilot study class.

One week was considered enough time for the pupils to prepare and have a short debate, as well as take the pre- and post-tests. The results of the pilot study gave clear indications of changes that needed to be made before the study was carried out on the research participants.

Firstly, the topic of the debate worked very well with the 7th grade class, they were engaged and enthusiastic. The only constructive feedback that was given was that they wanted more time to prepare for the debate and that the debate should have lasted longer. Both issues would naturally be solved in the full-scale experiment.

When it came to the tests, the feedback was more crucial. The pupils were given different lengths on their test to examine how long a test could potentially be. The feedback received showed that the tests were too long and boring for many of them. This offered valuable information that resulted in the tests being altered to be shorter for the research participants. Additionally, the pupils received different versions of the vocabulary tests to investigate whether the two versions were equal. Overall, the pupils achieved better scores on one of the tests, indicating that the test might not have been equal for Norwegian pupils their age. Due to this uncertainty regarding whether the two tests were in fact equal it was decided to only use one of the tests for the actual study.

Due to the small sample size and the different test lengths, all of the pupils were put in one group under the same conditions. Therefore, there was no information gleaned regarding the potential difference between the control and experimental group for the pilot study.

3.3 Participants

In the process of searching for participants for the actual study it was decided to reach out to schools the researcher was familiar with, while only having the criterion of it being a 7th grade

English class. This is called convenience sampling where one selects a sample based on the convenience of the researcher, such as availability and willingness to volunteer (Cohen, Manion, & Morrison, 2018, p. 218). 7th grade was chosen because of the competence aims from the Norwegian curriculum. They clearly state what the pupils should know when they exit 7th grade. It would, therefore, be expected that during the fall-term of 7th grade the pupils would already have competencies that cover most of the goals from the curriculum. The reason for only having that one criterion was due to the expected limited numbers of teachers in Norway willing to give away two and a half weeks of their time to a master student, especially during the pandemic.

After contacting several schools there were 74 pupils participating from two schools, with two classes participating from each school, in a city in the east of Norway. The classes were further divided in control- and experimental groups, resulting in four control- and four experimental groups. Altogether there were 30 participants in the control groups and 44 in the experimental groups. It was decided to keep the groups separated by class so that the pupils were in groups with the pupils from their regular class. There were similar distributions of characteristics in each group, determined by a questionnaire (section 3.5.3), to make the groups as equal as possible. A table of the different characteristics gathered from a questionnaire can be seen in the Appendix 1.

3.4 The lessons

The study took place over eight English lessons, with each lesson lasting forty-five minutes. This amounted to three lessons per week, plus two additional lessons used for testing. The control groups had lessons with their usual English teacher. The teachers were told to structure the control group lessons as they normally would for their own classes, with the added direction that they should refrain from working too extensively on vocabulary, or oral assignments. Since the teachers were familiar with the nature of the experiment, the last instruction ensured that they were mindful to adhere to their normal approaches, knowing it would be quite easy to influence outcomes unconsciously.

The experimental group had a two-week debate project. The topic for the debate was homework, which was chosen because it was relatable for the pupils and assumed to create interest and excitement amongst them. There were also several 7th-grade-friendly articles online about it, which made it more manageable to find useable sources. The preparations and the

actual debate used authentic English articles and conversation in English so it was not considered necessary to make the debate topic directly linked to the English curriculum.

The pupils in the experimental group were divided into four teams either being *for* or *against* homework in the debate panel, or *for* or *against* homework in the audience. In addition to that, they were assigned to play a role as either a parent, a teacher, or a pupil. One pupil could, for example, be a teacher in the audience who was against homework. The roles were implemented to encourage imagination and allow the pupils to step away from their own views (Schnurer & Snider, 2001, p. 66). There were two to four pupils on each team, depending on how big the experimental group was.

To prepare for the debate the pupils were given a booklet with articles for and/or against having homework during the first lesson. Since the booklet would form the basis of the pupils' arguments, they were given time during the lessons to read it. The booklet consisted of four longer articles and four tables/pictures/shorter list of information (Appendix 4). The longer articles in the booklet were all written for children and found on digital newspapers for children. Nevertheless, the newspapers were made in-and-for English speaking countries, so it was expected that some of the language would be a little challenging. Many of the articles had additional videos attached to them explaining much of the content in the articles. This gave the pupils the option to watch the video if it became too much to read all of the articles. They could also choose by themselves how many of the articles they wanted to read and if they wanted to supplement with additional sources they found online. The booklet contained some reliable and some less reliable sources, leaving it up to the pupils to decide which articles they would trust to form the basis of their arguments.

The lessons for the experimental group always started with talking about the topic of the day; then they worked with their team, or by themselves, on the topic (see Table 7). For example, during the fifth lesson, the term *rebuttal* was introduced. First, the importance of a good rebuttal was discussed. Then, the pupils worked together with their team to prepare rebuttals, where they used either the booklet or other digital sources to find arguments that they believed the other side would make. Lastly, they prepared a response to those arguments.

Table 7

Lesson Overview

Lesson	Treatment group
1	Pre-test
2	Introduction to debate and the booklet

3	Source reliability
4	Arguments
5	Rebuttals
6	Prepare for debate
7	Having the debate
8	Post-test

On the day of the debate, the classroom was set up to look like a debate room with the for and against team in front and the audience in the back. The for-homework side started with their arguments, then the against-homework side gave theirs. After that, the audience asked questions and the teams answered. The two debating teams were allowed to ask questions of each other and the audience was allowed to answer if the debating team could not. The researcher worked as the chairman deciding who could speak when. At the end of the lesson, a winner of the debate was chosen by secret ballot among participants. If the vote was tied, the chairman had the final decision about the winner.

3.5 Instruments

This study's findings are based on three types of material which are accounted for in this subchapter. In the following section the vocabulary test and attitude/anxiety survey will be presented, which were given as pre- and post-test to investigate dependent variables. Additionally, the questionnaire will be presented, which was conducted once to check for confounding variables, which are variables that might interfere with the findings, such as time spent playing English games, or months spent in an English-speaking country. The different tests/surveys took around 45 minutes all together. The tests were conducted the week before and after the treatment, resulting in a two-week gap between the pre- and post-tests.

3.5.1 The vocabulary test

As mentioned in section 2.1.4, the Updated Vocabulary Level test by Webb et al. (2017) was chosen to investigate the pupils' receptive vocabulary knowledge. This test was used both as a pre- and post-test to investigate the pupils' overall receptive vocabulary knowledge, and to look at the difference before and after the treatment. It was chosen because it was tested and validated by the authors in several L2 classrooms across the world and it was considered a very suitable

test for elementary learners of English as a foreign language (Webb et al., 2017; Victoria University of Wellington).

This vocabulary levels test is an updated version of The Vocabulary Level Test (VLT) created by Nation (1983). Webb et al. (2017) made a revised version that aimed to overcome two limitations identified in the previous versions. Firstly, the items used for the previous VLT were words that came from texts that originated in the 1930s and 1940s, which might not be relevant targets for today's pupils. Secondly, the VLTs did not measure the most frequent 1.000-word families, meaning, it started on level 2.000. According to Webb et al. (2017), it is very important to include 1.000-word families, since they account for as much as 80% of the words used in English (Webb et al., 2017, p. 34).

The Updated Vocabulary Levels Test (UVLT) works similarly to the VLT. It consists of ten three-item clusters per level and contains fifteen nouns, nine verbs and six adjectives per level. The pupils are supposed to tick off the word that they believe goes with the sentence (Table 8). The UVLT measures five levels, starting at level 1.000 (most frequent 1-1000 word families), 2.000 (most frequent 1001-2000 word families), and so on up, to level 5.000.

Table 8

Example of "The Updated Vocabulary Levels Test" by Webb, Sasao and Ballance (2017)

	game	island	mouth	movie	song	yard
land with water all around it		✓				
part of your body used for eating and talking			✓			
piece of music					✓	

Pupils can either take the whole test, or take individual levels (Webb et al., 2017, p. 33). For instance, pupils who cannot complete 2.000 or 3.000 levels should not be required to attempt 4.000, or 5.000-level sections. The value of the test is that it indicates which word frequency pupils should focus on during their learning. Therefore, the scoring of the test should focus on the individual levels not the overall score (Webb et al., 2017, p. 34). The pupils in this study were exposed only to levels 1.000-3.000, since that should be enough to cover the competence aims from the curriculum, hence, where the pupils and teacher should focus their attention (section 2.1.2.2). The test consists of two versions which are supposed to be used for a pre-test/post-test design. However, as mentioned in Section 3.2 the two versions did not yield equal scores during the pilot study. Therefore, the A-test was used for this thesis (see Appendix 1 in Webb et al., 2017).

As previously stated, the vocabulary test was given twice by the researcher, with a two-week interval between exposures. Before the pupils began the test, they were given instructions

that were suggested by the authors of the test. The pupils needed to provide the answers that they knew, or that they thought they knew, with the added direction that it was better to leave a question blank, than to guess randomly if they had no clue. If they wanted to hear a word read out loud, it was allowed. They could not, however, receive help with translating words, or with putting them into a context. Pupils were given the test during a 45-minute lesson but were allowed to sit longer with it if necessary. They were told that they were not meant to know every word and that this was not an easy test. In an effort to make the test a little less intimidating for them, they were informed that the aim was not to see how individuals did on the test, but how 7th graders performed on the test as a group. They were not told that they would retake the test in two weeks, but they were told that they would receive the test results after a while. When both tests were taken and assessed they received their total score on both the pre- and post-test. The teacher was given the test key, so that they could go through the test together.

3.5.2 Survey

As mentioned in section 2.3, vocabulary, L-attitudes and L-anxiety are closely related, but not a lot of research has been conducted on the relationship between them. Therefore, the pupils' L-attitudes and L-anxiety were investigated through parts of 'The Attitude/Motivation Test Battery' (AMTB) by Gardener (2004). This test or survey was made for secondary school students studying English as a foreign language, which is a little above the age that this study is investigating. However, previous studies have shown that it can be used with many different grades and with many different L1 speaking pupils. It has been tested on French-speaking students in Canada (Clément, Gardner, & Smythe, 1977), high school students in the Philippines (Gardner & Lambert, 1972), as well as, elementary school pupils in Estonia (Semerik, 2016). The Test Battery has also been validated and standardized for English-speaking Canadians, grades seven to eleven, to test their L-attitudes and motivations towards French. For this, researchers tested 1000 pupils from each grade (Gardner, 1985). Given these applications, it seems likely that the test could work in Norway with 7th grade pupils as well, with some modifications.

In the original AMTB there are eight categories with ten statements for each category. Each statement is followed by a Likert scale ranging from strongly disagree to strongly agree. For this thesis it was decided to only include three categories and four statements from each category, the survey used in this thesis can be seen in Appendix 3.

Research shows that pupils can answer one question very differently if the wording is only slightly varied. Converse and Presser (1986, p. 41) tested the sentence: "Do you think the United States should [forbid/not allow] public speeches against democracy?". Changing the word from *forbid* to *not allow* made a significant difference in the response of the participants. Therefore, Dörnyei (2010, pp. 24-25) suggests the use of multi-item scales, which are clusters of several differently worded items that focus on the same target. The item scores for the same target will then be added together, resulting in one total score. The idea is that one individual item might be misinterpreted, but then it will be averaged out during the summation of the item scores, reducing the damage of one individual item. Thus, it is advisable to have at least four items per subarea. If it ends up that one item is not comparable to the other items in the same subarea it can be excluded from the research (Dörnyei, 2010, p. 26).

The three subareas that were chosen for the questionnaire for this study were: 1) English class anxiety; 2) English use anxiety; and 3) Attitudes towards learning English. Both components for anxiety were added together to form the component, language anxiety. In the original AMTB there were ten statements for each of these subareas, so the four statements that seemed most suitable for 7th graders were chosen. In addition to the statements from the AMTB, one more statement was added to the post-test to investigate the L-attitudes, more specifically the enjoyment the pupils had in regard to the two-week treatment.

The four statements for each subarea were checked using Cronbach's alpha test in SPSS. All of the subareas had a Cronbach's alpha above .7 on both the pre- and post-test. Which, according to Dörnyei (2007, p. 207), is acceptable because the reliability coefficients should be at least .7 in L2 research. This implies that all of the subareas are within an acceptable area to consider that all of the four statements within each subarea represent the same construct and can, subsequently, be grouped together and represent one main score. It is debatable whether summed scores like these should be looked at as ordinal or interval scales for the purpose of statistical analyses. While interval data can be measured with mean, median etc. ordinal data cannot. According to Brown (2011) there is a difference between Likert items and Likert scales. Likert items are each individual item in a survey and should be looked at as ordinal data. Likert scales on the other hand, are a total score of multiple Likert items and can be looked at as interval data as long as the items are checked for reliability with a measurement like Cronbach's Alpha. Therefore, the L-attitude and L-anxiety scores will be treated as interval data for this study.

The survey was translated into Norwegian so that there would not be any language barriers. The Likert scale had six response options ranging from strongly disagree to strongly

agree. The neutral (neither agree nor disagree) option was omitted because it has not seemed to have an effect in previous studies (Dörnyei, 2010; Johns, 2005) and to avoid the pupils choosing the middle value because they became too tired or bored to make a choice.

3.5.3 Questionnaire

The questionnaire was made to provide general background information about the participants and potentially find confounding variables that could influence outcomes on the tests. Potential confounding variables that were checked for were, for example: time spend in an English-speaking country, language spoken at home, or time spent engaging in English spare time activities (games, TV, books etc.) (see questionnaire in Appendix 1). This was mainly done to have as equal groups as possible and to investigate other potential reasons to obtain certain scores. If, for example, one group scored highly on the vocabulary test, and that group had a majority of pupils speaking English at home, then that could be an explanation for the high score.

The criteria for this questionnaire were that it needed to be, or be inspired by, an already tested questionnaire, given the lack of time to have a full-size pilot study, as well as ensure that the questionnaire could provide reliable answers. Additionally, it had to be suitable for 7th graders. The questionnaire was made with inspiration from a questionnaire Sundqvist (2009) used in her dissertation on 80 Swedish 9th graders. Her study investigated similar aspects and similar aged pupils, so it was considered suitable for this study as well. Several of the questions used in this study were identical or similar to Sundqvist's questionnaire. However, some changes were made to strive for anonymity with the pupils; for example, instead of asking what language they spoke at home, they were asked whether they spoke Norwegian, English or other (question 2, Appendix 1). The questionnaire was also translated to Norwegian and made shorter so the pupils would not get bored.

The questionnaire was administered by the researcher along with the pre-tests on the first day. Before starting, the researcher read through the questions and explained some of the trickier examples. The pupils were also allowed to ask for further clarifications while answering the questionnaire if something was unclear.

None of the pupils were excluded from the study based on the answers in the questionnaire, since the two groups had very similar characteristics (Appendix 2). Due to time limitations, the questionnaire has not been used to look for correlations to the other tests.

3.6 Statistical methods

The data from the measuring instruments were computed using IBM SPSS version 27. The data was first recorded into excel and then transferred to SPSS. Different statistical tests were run to answer the different research questions. This section will quickly go through the different tests. The results from the tests for this thesis are presented in chapter 4.

3.6.1 Data preparation

The first step after the tests and surveys were completed was to code the responses. The vocabulary test provided numbered scores, so those tests had to be scored manually. However, the participants' answers for the survey and questionnaire were not numbered scores, and, therefore, had to be converted to numbers by means of "coding procedures" (Dörnyei, 2007, p. 199). The coding procedure involves defining each variable and then making coding specifications for every possible value in a variable. For example, for the Likert items from the survey, the values were ranked from 1-6 (1=strongly disagree, 2=disagree, 3=partly disagree, 4=partly agree, 5=agree, 6=strongly agree). For the questionnaire each possible answer received a specific value, for example, gender got three values (1=boy, 2=girl, 3=do not want to answer). Since the data only included close-ended questions, there were limited numbers of possible answers which made the coding process easier. During the coding of the survey, the negatively worded values were reversed manually.

After the coding procedures, the values had to be inserted into the statistical software. The data was first keyed into excel and then transferred to SPSS. In SPSS the missing values were set to 99, and the codes for the data were implemented. The data was cleaned and screened by double checking outliers and every case that had very big differences from pre- to post-tests.

3.6.2 Outliers and missing data

There are several ways to deal with outliers and missing data, Larson-Hall (2016, p. 38 & 124) points out that how to deal with both outliers and missing data are rich questions that do not have a single and straightforward answer. The researcher has to make a decision on how to best deal with it for the specific study. For this study, the outlying scores were checked again to see if any errors were made while calculating the scores. Four participants were removed from the vocabulary test scores because of discrepancies in their scores. Their tests showed signs of them randomly ticking off boxes or just stopping in the middle of one of the tests and not the other. These participants were, therefore, removed, but only from the vocabulary scores and not from

the L-attitude/L-anxiety survey. In that way their inconsistencies would only affect the area in which they showed inconsistencies and it was considered the best option on how to deal with these participants. Other outliers that were found did not show signs of ambiguous answers and several analyses were conducted with and without the outliers, and the outcomes showed very similar results, so the other outliers were kept in the data.

There was no missing data from the vocabulary test, but for the survey several participants had missing data. The missing data for the survey seemed to be mostly due to the pupils skipping a statement or answering twice for one statement while skipping the next. These missing values seemed random and not intentional. Therefore, nothing was done to these participants in the descriptive statistics, and in the inferential statistics the cases were excluded test-by-test or pairwise, which is the default in SPSS. According to Larson-Hall (2016, p. 38), this is not considered the best way to deal with missing values, but given the few participants this affected, it was considered the most efficient and comprehensible way to deal with for a non-statistician.

Once the data was coded, inserted into SPSS, screened, and cleaned, the statistical analysis could be run.

3.6.3 Shapiro Wilk's test

The Shapiro Wilk's test is a formal statistical test to check the data for normal distribution. If the significance level is below .05 the null hypothesis is rejected, and there is reason to believe that the data is not normally distributed. Subsequently, if the data is not normally distributed, nonparametric tests need to be used when assessing the data, and if the data is normally distributed, parametrical tests can be run (Norušis, 2008, p. 266).

3.6.4 Paired-Samples T-test and Wilcoxon Signed-Rank test

The Paired-Samples T-test and the Wilcoxon Signed-Rank test are both tests that compare two sets of scores obtained from the same group, or, as in this thesis, when the same participants are measured more than once (Dörnyei, 2007, p. 115). The Wilcoxon Signed-Rank test is a nonparametric test that can be used instead of the parametric Paired-Samples T-test. These tests can tell us if the difference between two scores from the same participant is big enough to reach statistical significance ($p < .05$). Hence, if there is a statistically significant increase or decrease in scores from the pre- to the post-test for vocabulary, L-attitudes and L-anxiety.

3.6.5 Independent-Samples T-test and Mann-Whitney U test

The Independent-Samples T-test and Mann-Whitney U test are both tests that compare the difference in scores between independent groups (Dörnyei, 2007, p. 115). The Independent-Samples T-test in the parametric test, whereas the Mann-Whitney U test is the nonparametric alternative. For this thesis, these two tests investigate two things, firstly, whether there were significant pre-existing differences between the groups by testing the pre-test scores (i.e., if $p < .05$). Secondly, if the groups received significantly different results for the post-test (i.e., if $p < .05$). If the pupils had no significant pre-existing differences from the pre-test and then received a significantly different score on the post-test, it would indicate that one group improved more than the other after the treatment period. However, if the groups had significant pre-existing differences from the beginning, it would be a confounding variable which could influence the results from the post-test.

3.6.6 Spearman's Rho test

The Spearman's Rho test is the nonparametric alternative to the Pearson correlation test. These tests examine the correlation/relationship between variables by evaluating the strength and direction of this relationship. The correlation coefficient ranges from -1 to 1, including fractional possibilities. A high coefficient implies a strong positive relationship, which indicates that if the pupils get a high score on one variable, they should also get a high score on the other variable. A coefficient of 0 suggest no relationship between the variables, and a negative coefficient suggest inverse relationships where one score is high when the other is low. The correlation coefficient also needs to be statistically significant to indicate a true correlation (Dörnyei, 2007, p. 223). For this thesis the relationship between vocabulary, L-attitudes and L-anxiety are investigated.

3.7 Ethical considerations

The participants in this study were minors, which presents several ethical aspects that need to be taken into consideration. A notification form for personal data was sent to NSD (Norwegian center of research data) to ensure that the project followed the ethical guidelines. Once NSD approved the project (Appendix 6), an information sheet was sent to the teachers along with a consent form that needed to be signed by the pupil's guardians (Appendix 5). The fact that participation was voluntary was stressed to the teachers, pupils and parents. The pupils were told that they could decide to not participate even if their parents had signed the consent form.

They had the freedom to withdraw their participation at any point and if they did not want to participate, it would not affect them in any way. They would then follow the lessons of the control group and they would read their book while the others took the tests.

Parents received several emails from the teachers reminding them to hand in the consent forms, but because of time constraints, not all consent forms were handed in in time. Pupils without consent forms did not participate in the study.

It was also important to ensure the pupils' anonymity. Therefore, the tests were all taken using pen and paper, and the pupils were given a unique ID-number that only they knew. This was also done to make the study blind so the data would be as valid as possible. The ID-number was used to be able to compare the questionnaire, pre- and post-test scores of the pupils. Even though there was no way to know which test belonged to which pupil, there was a chance that the questionnaire could potentially give away some pupils: for example, only one pupil has lived abroad for more than 6 months (question 5, Appendix 1) and also speaks another language at home (question 2, Appendix 1). This was explained to the pupils and informed about in the consent form to the parents. It was also stressed that this information would not be published in the final thesis and that their children would be part of a statistic and not presented individually in the publication.

3.8 Validity and reliability

Validity and reliability function as a sort of quality criteria for studies. There are several different types of validity and reliability and they have different definitions throughout literature. Common for all of the definitions is the fact that threats to validity and reliability can never be completely erased and there is no such thing as perfect validity or reliability. However, the threats to validity and reliability can be reduced by giving it attention throughout the research and by discussing how these threats might affect the study. This section will, therefore, discuss different threats in this study to validity and reliability.

3.8.1 Validity

Validity has to do with how valid the results of the study are. The results should reflect what they are believed to reflect. Additionally, the results need to be meaningful not only to the sample tested, but also to the larger relevant population (Mackey & Gass, 2015, p. 158). There are many types of validity, and for this chapter it has been divided into measuring instruments, participants and overall generalizability.

3.8.1.1 Measuring instruments

The first potential threat to validity has to do with whether or not the measuring instruments are measuring what they are supposed to measure. The vocabulary test and the attitude/anxiety survey have earlier been used in research in several L2 classrooms across the world, which indicates that they produce valid data. Yet, most studies are slightly different and strive to measure slightly different aspects, and, therefore, the threats to validity need to be considered when using the measuring instruments in this particular study (Bachman, 2004, p. 260).

Firstly, the vocabulary test is one out of many ways to test receptive vocabulary knowledge. The test provides information about the pupils' vocabulary level which is one way to measure vocabulary knowledge. As discussed in section 2.1.4, vocabulary knowledge can also be measured using size tests, and potentially the best measurement of vocabulary knowledge would have been to have both a size test and a levels test. This would, however, have created other threats to validity such as boredom and inattention. Consequently, the decision was made to only use one test and referring to the discussion in section 2.1.2 on the importance of knowing the learners' vocabulary level, UVLT was chosen to represent the construct, receptive vocabulary knowledge.

Another choice was made to only measure the first three-thousand levels. The decision to include the first three levels was due to the importance of learners to learn the high-frequency words and level 3.000 (see section 2.1.2). Nevertheless, this decision could hinder the possibility to find out every pupils' full potential when it comes to vocabulary knowledge. Regardless, it was considered the best option while still limiting boredom and inattention.

When it comes to the L-attitude and L-anxiety measuring instruments, the scores cannot say anything about the actual L-attitude and L-anxiety of the participants, but rather about their self-reported L-attitudes and L-anxiety. Consequently, the research questions ask specifically for the self-reported L-attitudes and L-anxiety and not their actual attitude and anxiety.

The use of multi-item scales for these measurements have made the constructs more valid since there are four to eight items that make up the constructs, and it is, therefore, more likely that the scores actually reflect the pupils' self-reported L-attitudes and L-anxiety (Dörnyei, 2007, p. 104). In that way it is more likely that the score actually reflects the participants beliefs regarding the construct and not just misunderstandings of a statement.

Additional issues with the measuring instrument are known as *test effect* and *practice effect*. Practice effect has to do with improved test scores, due to the participants gaining more experience in taking that particular test (Dörnyei, 2007, p. 53). Test effects has to do with the

equivalence between the pre- and post-test (Mackey & Gass, 2015, p. 168). This is mainly an issue with the vocabulary test. The practice effect should be limited due to the thorough explanation of the test, and the assistance the pupils could receive throughout the test. The test effect on the other hand, poses a potential threat to validity. The UVLT consists of two equal tests that are supposed to function as a pre- and post-test set. Yet, to my knowledge, they have not been tested in the Norwegian classroom. Hence, resulting in an unknown equivalent between the two tests in a Norwegian setting. Because of this, there was a risk of getting a test effect scenario, where one of the tests are easier than the other, which could yield unreliable test scores. If the pre-test was easier, the pupils would have scored worse on the post-test regardless of their actual progress, and vice versa. The possibility of one of the tests being easier than the other was heightened when the participants in the pilot received an overall better score from one of the tests (see section 3.2). Therefore, the choice was made to administer the same test twice. This, however, leads to other potential issues where the pupils might have investigated the meaning of certain words before the post-test. Considering that the test consisted of 90 words and 150 potential answers that could be matched with those words, it was considered less likely that they would remember words to search for at home, and more likely that the two tests would have been unequal. This could potentially still affect the results of the test in some other way that has not been considered.

3.8.1.2 Participants

A second threat to the validity of the data is the participants. This has to do with how the pupils answered the tests/surveys and threats that would influence their answers.

As mentioned throughout the last paragraph, several decisions were made to limit the threats of participation boredom, fatigue and inattention. Due to the feedback from the pilot test, the vocabulary test was shortened to only include the first three-thousand levels. The issue of boredom and fatigue could potentially lead to invalid results if multiple pupils stopped halfway through the tests because they did not have the energy or will to finish the test and do their best. To further limit this threat the tests could have been even shorter, but this again would have had implications for the validity of the measuring instruments in terms of the extent to which they represent the constructs. Hence, if the vocabulary test was shortened to only contain 10 items each level instead of 30, it would have limited the potential for boredom further, but it would not have captured the levels adequately. The same goes for the L-attitude and L-anxiety scores, Dörnyei (2007, p. 104) argues that four to ten items should be included to sufficiently capture a construct.

Other threats were considered such as the Hawthorn- and Halo effect, which are both situations where the participants answer the tests and surveys in a different way because they are part of an experiment or their teachers and researcher are present (Mackey & Gass, 2015, p. 166). To limit this, the participants were told multiple times that no one would know which test belonged to them, hence their anonymous ID-number (Section 3.3), and that the point of the study was to get an overview of 7th graders responses, and not individual pupils' responses. This way the pupils did not have to feel pressured to perform in a certain way, although it is not possible to be sure that they did not feel pressured.

The physical environment and setting where the participants took the test can also influence the results. Therefore, both the control groups and experimental groups took the tests in the same classroom at the same time, so that factors like classroom noise and taking the test at the end of the day would affect both groups similarly. The pre- and post-tests were also administered around the same time during the day for each group so that the pupils would have approximately the same energy level.

3.8.1.3 Generalizability of the sample

A third threat is the generalizability of the findings, which concerns whether the findings are relevant to the wider population or just to the sample. A quasi-experimental study, as this study, strengthens the generalizability due to the use of intact classes and authentic learning (Dörnyei, 2007, p. 120). Since the research takes place in a natural learning environment with intact classes it is easier to transfer the study to usual lessons. A change in learning environment and classes could potentially become a confounding variable, which would interfere with the results. Not using intact classes would also have been a logistical issue since the classes have different schedules. At the same time, the use of intact classes dismisses the possibility of having a random sample of participants. As mentioned in section 3.3 the participants were chosen based on convenience sampling, which suggests that the participants all come from the same area, have had the same teachers, etc. which can affect the generalizability. According to Mackey and Gass (2015, p. 172), the generalizability of the study is strengthened when the sample is drawn randomly from the population. This suggests that every individual has the same opportunity as being selected as another individual. However, this is difficult to achieve, and therefore, most research in applied linguistics use non-random sampling (Dörnyei, 2007, p. 98). The use of a non-random sample could potentially lead to results that are due to pre-existing differences and not the treatment for the quasi-experimental part of the study. However, this

can be, and was, checked for in this thesis, and no significant pre-existing differences were found between the groups on either test.

It can also be the case that the four classes that participated have abnormally high or low vocabulary knowledge. This could have been avoided if the sample was much larger and randomly assigned, but that would have demanded much more time and resources, which was not available for this study, and is usually an unrealistic and not feasible aim for L2 research in general (Dörnyei, 2007, p. 98). Additionally, including such a large and randomly assigned sample would have interfered with the quasi-experimental aspect of the study, which for practical reasons needed to have intact classes. It was, therefore, considered that it was worth keeping the intact classes and authentic learning environment and risk having the non-random sample.

3.8.2 Reliability

Reliability refers to the dependability, consistency and replicability of certain instruments in certain groups. It has to do with whether the results are believable, and if similar results would be found if the study was conducted on a similar group of participants in a similar context at a different time (Cohen et al., 2018, p. 268).

Cohen et al. (2018, p. 268) claims that a reliable instrument would yield similar data from similar respondents. This implies that for this study the control group and experimental group should have obtained similar scores on the pre-tests due to their similarities in characteristics. In this study, the participants from both groups had similar scores for the pre-tests, where no significant differences were found between the groups of neither the vocabulary pre-test nor the survey ($p > .05$).

Another way to ensure reliability within the study is to check the consistency between the pre- and post-scores. For this thesis the results from the pre- and post-tests were tested through the Pearsons correlation test to make sure that there was sufficient correlation between the two administrations of the tests. All of the tests showed a statistically significant correlation between the pre- and post-test (the vocabulary tests ($p < .001$), the attitude tests ($p = .012$) and the anxiety tests ($p < .001$)). This indicates that the tests produced consistent results for this study. The inclusion of several items to form the constructs L-attitude and L-anxiety was also done to strengthen the reliability. The items checked for correlation to form the multi-item scales and the Cronbach's Alpha test showed that the multi-item scales for both L-anxiety (.805) and L-anxiety (.848) had high internal reliability.

Additionally, the fact that the results of the tests and surveys for this thesis produce numbers and scores strengthens the reliability since there is not much interpretation and judgement involved in the analysis of the data, in comparison to qualitative research using, for example, semi-structured interviews. The data from this study was simply inserted into excel, and to make sure that the numbers were added correctly, a selection of the numbers were double-checked.

4 Results

This chapter will present both the descriptive and inferential results of this study to be able to further discuss the research questions in chapter 6. The descriptive statistics will report the number of participants, a measure of central tendency (e.g., mean, median) and a measure of spread (e.g., standard deviation, range). The inferential statistics will present statistics which determines the generalizability of the results from the sample. This chapter is divided up into four sections.

Firstly, the results from the vocabulary tests will be presented, which will help shed light on research questions one and four: The first research question, "What is the receptive vocabulary knowledge of Norwegian 7th graders?", will be viewed through the descriptive statistics from the pre-test. These statistics will indicate the vocabulary knowledge of the sample. The fourth research question, "Does engaging the learners in a collaborative output task in the form of a classroom debate have an effect on the pupils' receptive vocabulary knowledge?", will be viewed through the descriptive and inferential statistics of vocabulary knowledge both within the groups and between the groups.

Secondly, the results from the L-attitudes scores will be presented, followed by the L-anxiety scores, which will both help shed light on research questions two and five. Research questions two, "What are the pupils' attitudes towards learning English and their levels of anxiety when using English?", will be viewed through the descriptive statistics from the pre-test. These statistics will give indications on the L-attitudes and L-anxiety of the sample. Research question five, "Does engaging the learners in a collaborative output task in the form of a classroom debate have an effect on the pupils' L-attitudes and L-anxiety?", will be viewed through the descriptive and inferential statistics of L-attitudes and L-anxiety both within the groups and between the groups.

Lastly, correlational statistics between vocabulary level, L-attitudes and L-anxiety will be presented to help answer research question three "Does engaging the pupils in a collaborative output task in the form of a classroom debate have an effect on the pupils' language learning attitudes and language anxiety?"

4.1 Receptive vocabulary knowledge

This section will be divided up into two main sections according to research questions one and four. The first section will focus on research question one and present descriptive statistics for the sample. The second section will focus on research question four, and present both

descriptive and inferential statistics to shed light on the effect that a collaborative output task, such as debate, has on receptive vocabulary knowledge.

4.1.1 The receptive vocabulary knowledge of the sample

To address the first research question, the receptive vocabulary knowledge of the participants needs to be investigated. The UVLT was administered both as a pre- and post-test. For this section only the pre-test scores will be used to address the research question. The results provided four scores from each test, namely scores for level 1.000, 2.000, 3.000 and a total score for all levels. The maximum score within each of the thousand levels was 30 points. As seen in Table 9, the participants (N = 70) had the highest average score with the smallest standard deviation for level 1.000 (M = 27.40, SD = 3.24), followed by level 2.000 (M = 21.09, SD = 6.57), and lastly, level 3.000, with the lowest average score and the largest standard deviation (M = 15.87, SD = 6.67). The average total score for all levels was 64.36 (SD = 15.39). Some pupils got the maximum score of 30 within each level, and the minimum score across the levels was 1, which was obtained at level 3.000.

Figure 3 illustrates the distribution of variables visually with a box plot, where the grey box illustrates the interquartile range, the lower boundary of the box represents the 25th percentile and the upper boundary represents the 75th percentile. This implies that 50% of all scores lie within the grey box. The line inside the box represents the median. The "whiskers" extend to the largest and smallest observed values within 1.5 box lengths, and the dots outside of the whiskers are outliers. Hence, this box plot illustrates both the median and spread of the data. For example, the box representing level 1.000 has the median line closer to the top of the box, which suggest that there is a tail towards smaller values, where the length of the tail is shown by the length of the whiskers and outlying values (Norušis, 2008, p. 126).

Table 9, along with the data frequency, gives information about the level of mastery of the different frequency levels. Even through there is not one universal threshold for when full mastery is achieved, Webb et al. (2017, p. 56) suggested to put the threshold for mastery at 29/30 for the first three-thousand levels on their UVLT. They mention that such a high threshold is necessary for those frequency-levels since they account for such a large portion of running words. The leaners should, therefore, achieve near perfect knowledge of the words in a level, before moving on to the next. However, Webb et al. (2017, p. 56) put the threshold at 24/30 for the levels above 3.000, since they account for a much smaller percentage of running words.

According to the threshold of 29/30 points, 51.4% (36/70) of the participants mastered level 1.000, while 17.2% (12/70) mastered level 2,000 and 5.8% (4/70) mastered level 3.000.

The results show that the sample acquired a relative high mastery of level 1.000, and low mastery of level 3.000.

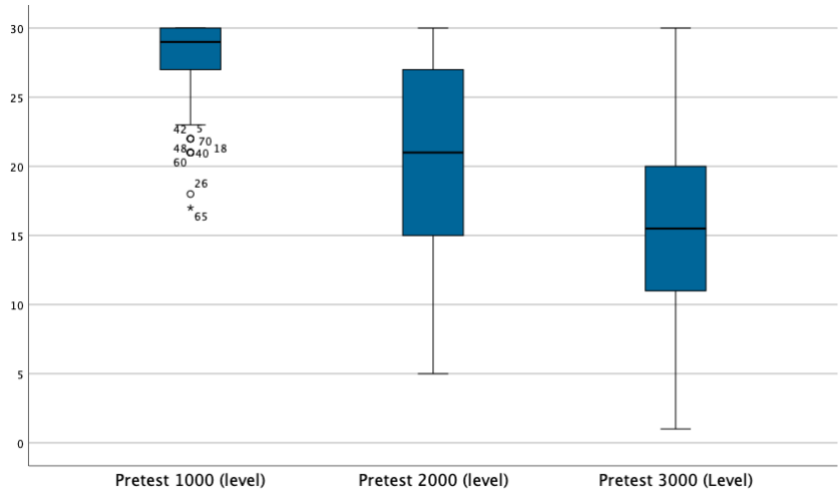
Table 9

Descriptive Statistics of Vocabulary Pre-test Scores

	N	Minimum	Maximum	Mean	Std. Deviation
Pretest 1000 (level)	70	17	30	27.40	3.237
Pretest 2000 (level)	70	5	30	21.09	6.565
Pretest 3000 (Level)	70	1	30	15.87	6.668
Pretest Vocabulary	70	24	90	64.36	15.387
Valid N (listwise)	70				

Figure 3

Box plot of Vocabulary Pre-test Scores



4.1.2 The effect of a collaborative output task (classroom debate) on the pupils' receptive vocabulary knowledge

To begin to answer research question four, two main aspects must be looked at. Firstly, the difference between pre- and post-test scores for vocabulary levels within the groups will be presented. Secondly, the difference in scores between the two groups will be viewed. Before the tests can be run, the groups need to be checked for normal distribution. As seen in Table 10, both groups were normally distributed, as assessed by the Shapiro-Wilk's test ($p > .05$), except for the control group post-test, which is just barely below the threshold. Considering that the other three scores were normally distributed, it is considered acceptable to use parametrical tests, namely the Paired-Samples T-test and the Independent-Samples T-test for these scores.

Table 10
Shapiro-Wilk's Test Vocabulary

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
Group		Statistic	df	Sig.	Statistic	df	Sig.
1 Control group	Pretest Vocabulary	.091	27	.200*	.978	27	.811
2 Experimental group	Pretest Vocabulary	.106	43	.200*	.965	43	.212
1 Control group	Posttest Vocabulary	.126	27	.200	.924	27	.049
2 Experimental group	Posttest Vocabulary	.102	43	.200*	.950	43	.061

4.1.2.1 Vocabulary pre- and post-test scores – Within groups

A Paired-Samples T-test was conducted to determine whether the pupils' vocabulary scores were statistically different from the pre-test to the post-test. The control group (N = 27) received a higher score on the post-test (M = 65.89, SD = 16.04) as opposed to the pre-test (M = 63.22, SD = 14.71) (Illustrated in Figure 4). Six pupils had a decrease in their scores from the pre-test to the post-test, 19 pupils had an increase and two had the same score on both tests. According to the Paired-Samples T-test, which can be seen in Table 11, there is a statistically significant mean increase in vocabulary scores of 2.667, 95% CI [.867, 4.466], t (26) = 3.046, p = .005, d = .59.

The experimental group (N = 43) received a higher score on the post-test (M = 67.51, SD = 15.74) as opposed to the pre-test (M = 65.07, SD = 15.93) (Illustrated in Figure 4). Eight pupils saw a decrease in scores from the pre-test to the post-test, 30 pupils had an increase, and five had the same score. According to the Paired-Samples T-test, which can be seen in Table 11, there is a statistically significant mean increase of 2.442, 95% CI [1.203, 3.618], t (42) = 3.978, p < .001, d = .61. Both groups had a statistically significant increase in scores from the pre-test to the post-test. The data from both of the groups is illustrated visually in Figure 4, the median line increases for the post-test for both groups.

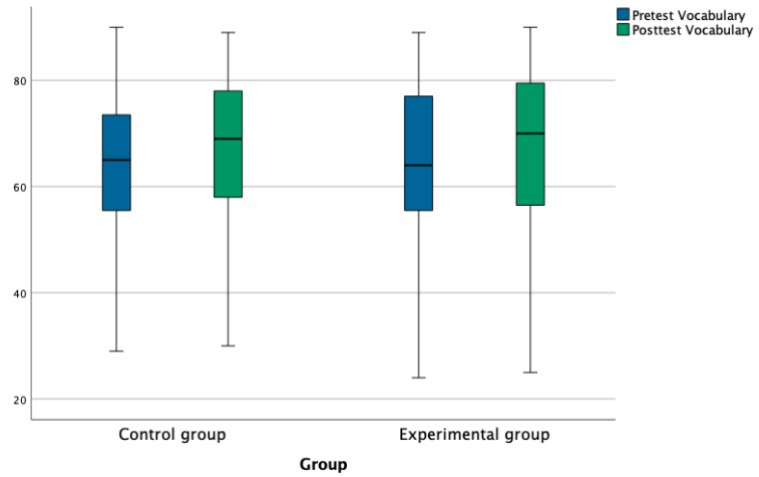
Table 11

Paired-Samples T-test vocabulary

			Paired Differences							
						95% Confidence Interval of the Difference				
Group			Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
1 Control group	Pair 1	Posttest Vocabulary - Pretest Vocabulary	2.667	4.549	.875	.867	4.466	3.046	26	.005
2 Experimental group	Pair 1	Posttest Vocabulary - Pretest Vocabulary	2.442	4.026	.614	1.203	3.681	3.978	42	.000

Figure 4

Box plot Vocabulary Within Groups



4.1.2.2 Vocabulary pre- and post-test scores – between groups

An independent-samples t-test was run to investigate whether there was a difference in the vocabulary test scores between the control group (N = 27) and the experimental group (N = 43). To do this, separate t-tests were run for the pre- and post-test scores. The pre-test was used to check if there were any preexisting differences between the two groups, and the post-test was used to see if there was a difference between the two groups after the treatment. There was homogeneity of variance for both the pre-test (p = .415) and post-test (p = .865), as assessed by Levene's test for equality of variance. For the pre-test scores, the experimental group (M = 65.07, SD = 15.93) had a higher average score than the control group (M = 63.22, SD = 14.71), but the difference was not statistically significant, M =1.85, 95% CI [- 5.734, 9.429], t(68) = .486, p = .628, as seen in Table 12.

Similarly, for the post-test scores, the experimental group (M = 67.51, SD = 15.74) had a higher average score than the control group (M = 65.89, SD = 16.04), but the difference was not statistically significant for the post-test either, M = 1.623, 95% CI [- 6.14, 9.34], t(68) = .865, p = .678. The results from the Independent-samples T-test can be seen in Table 12. The scores from the pre- and post-tests can be seen in Figure 5. The blue boxes illustrate the pre-test scores for both groups, while the green boxes illustrate the post-test scores.

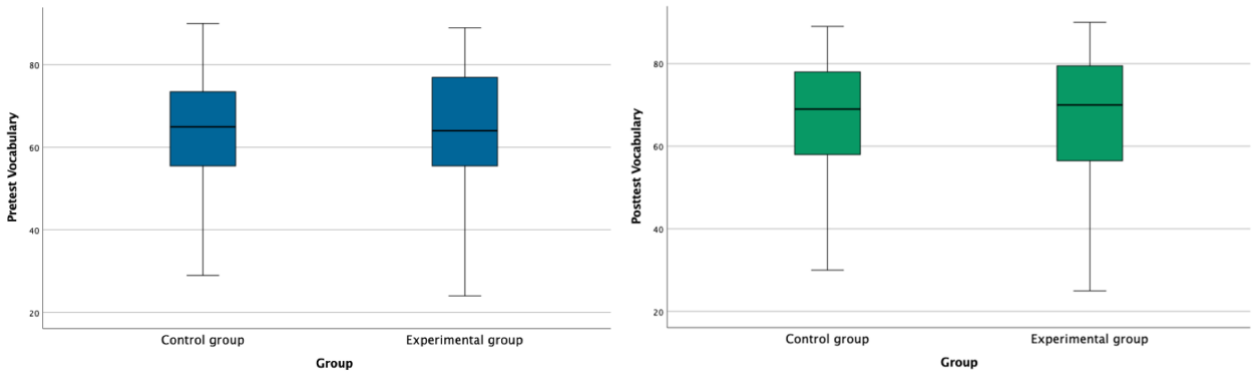
Table 12

Independent-samples T-test Vocabulary Levels

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Pretest Vocabulary	Equal variances assumed	.673	.415	.486	68	.628	1.848	3.799	-5.734	9.429
	Equal variances not assumed			.495	58.666	.622	1.848	3.730	-5.618	9.313
Posttest Vocabulary	Equal variances assumed	.029	.865	.417	68	.678	1.623	3.892	-6.144	9.390
	Equal variances not assumed			.415	54.580	.680	1.623	3.910	-6.214	9.459

Figure 5

Box plot Vocabulary Between Groups



4.2 Language attitude

This section will be divided up into two main sections according to research questions two and five. But firstly, the items making up the multi-item scale L-attitudes must be tested for internal reliability to check if they can be added together. The Cronbach's alpha test was run to test for internal reliability, which showed that the multi-item scales for L-attitudes have high internal reliability (.805). Consequently, the item scores could be added together to form the multi-item scales L-attitudes.

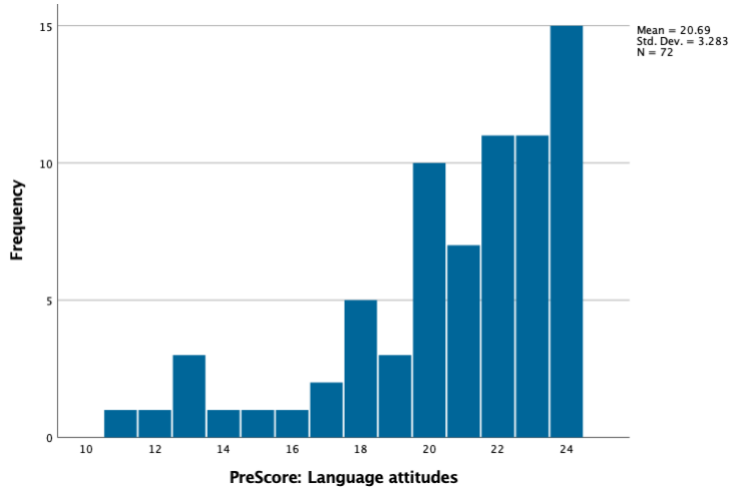
4.2.1 The language attitudes of the sample

To comment on research question two, the language attitude scores of the sample need to be investigated. The multi-item scale L-attitudes consists of four item scores. The total score has a minimum score of 4 and a maximum score of 24, where a high score indicates positive attitudes towards learning English. The L-attitude scores show that the participants (N = 72) have a minimum score of 11, a maximum score of 24, and a high average score (M = 20.69, SD

= 3.283). This indicates that more pupils have positive than negative attitudes towards language learning. This is visible in Figure 6, where the scores for L-attitudes are skewed to the right, indicating a cluster of positive attitudes. However, it is also visible that some participants reported rather negative attitudes towards learning.

Figure 6

Histogram Language Attitudes Scores



4.2.2 The effect of a collaborative output task (classroom debate) on the pupils' attitudes towards learning English

To comment on research question five, two main aspects must be looked at. Firstly, the difference between pre- and post-test scores for language attitude within the groups will be presented. Followed by the difference in scores between the two groups. Before the tests can be run the groups need to be checked for normal distribution. The results from the normal distribution test are displayed in Table 13 and it indicates that neither of the groups, for neither the pre- nor post-test, were normally distributed, as assessed by the Shapiro-Wilk's test (p < .05). Subsequently, the inferential statistics need to be obtained with the use of non-parametric test, namely, Wilcoxon signed ranks test and Mann-Whitney U-test.

Table 13

Shapiro-Wilk's test for Language Attitudes

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
Group		Statistic	df	Sig.	Statistic	df	Sig.
1 Control group	PreScore: Attitudes towards learning English	.222	29	.001	.848	29	.001
2 Experimental group	PreScore: Attitudes towards learning English	.183	43	.001	.859	43	.000
1 Control group	PostScore: Attitudes towards learning English	.186	30	.009	.878	30	.003
2 Experimental group	PostScore: Attitudes towards learning English	.163	41	.008	.883	41	.001

a. Lilliefors Significance Correction

4.2.2.1 Language attitude pre- and post-test scores – Within groups

The Wilcoxon signed ranks test was conducted to determine whether the pupils' L-attitudes were statistically different from the pre-test to the post-test. In the control group (N = 29), nine pupils reported having less positive attitudes towards learning English on the post-test as opposed to the pre-test, 13 reported more positive attitudes, and seven reported the same attitudes on both tests. There was a median decrease of 0.5 points in reported L-attitudes from pre-test (*Mdn* = 22.0) to post-test (*Mdn* = 21.5), but this was not statistically significant, z = -.592, p = .554 (Table 14).

From the experimental group (N = 40), 16 pupils reported having less positive L-attitudes towards learning English on the post-test as opposed to the pre-test, 15 reported more positive attitudes, and nine reported the same L-attitudes on both tests. There was a median increase of 1 point in reported L-attitudes from pre-test (*Mdn* = 21.0) to post-test (*Mdn* = 22.0), but this was not statistically significant, z = -.337, p = .736 (Table 14). Figure 7 illustrates the participants differences in scores from the pre-test to the post-test. The blue boxes illustrate the pre-test scores and the green boxes illustrate the post-test scores.

Table 14

Wilcoxon Signed Ranks Test Language Attitude

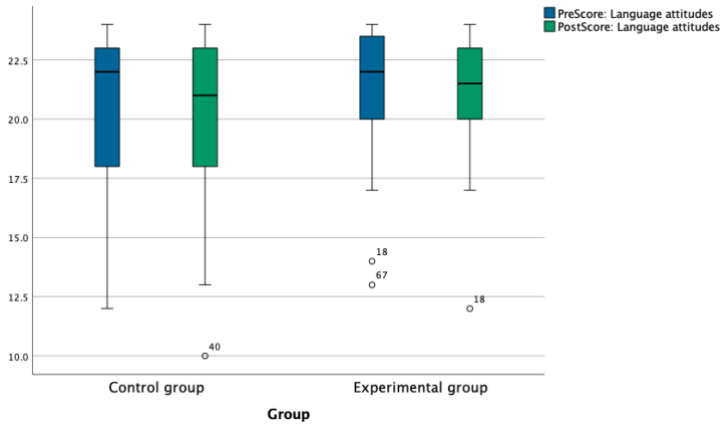
Group		PreScore: Language attitudes - PostScore: Language attitudes
1 Control group	Z	-.592[b]
	Asymp. Sig. (2-tailed)	.554
2 Experimental group	Z	-.337[b]
	Asymp. Sig. (2-tailed)	.736

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

Figure 7

Box plot Language Attitudes Within Groups



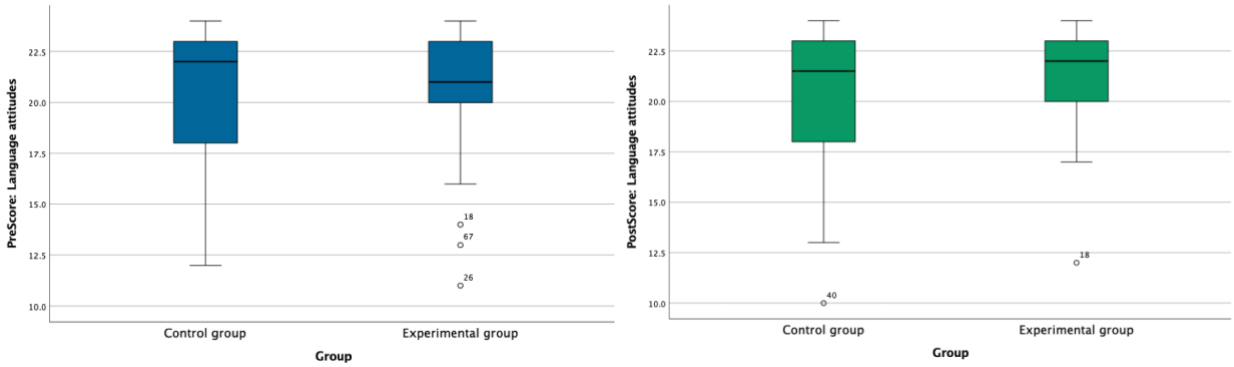
4.2.2.2 Language attitude pre- and post-test scores – between groups

A Mann-Whitney U test was run to determine if there was a significant difference in language attitude scores between the control group and experimental group for both pre- and post-test. Distributions of the language scores were similar, as assessed by visual inspection. For the pre-test the control group (*Mdn* = 22.0, *N* = 29) received a higher median than the experimental group (*Mdn* = 21.0, *N* = 43), but the difference was not statistically significant, *U* = 654, *z* = .354, *p* = .724.

For the post-test the control group (*Mdn* = 21.5, *N* = 30) received a lower median than the experimental group (*Mdn* = 22.0, *N* = 41), but the difference was not statistically significant, *U* = 662, *z* = .553, *p* = .580. Both the scores from the pre- and post-test are illustrated with a box plots in Figure 8 where it is visable that the scores from the experimental group clusted more towards the higher values than the control group scores. The blue boxes illustrate the pre-test scores for both groups, while the green boxes illustrate the post-test scores.

Figure 8

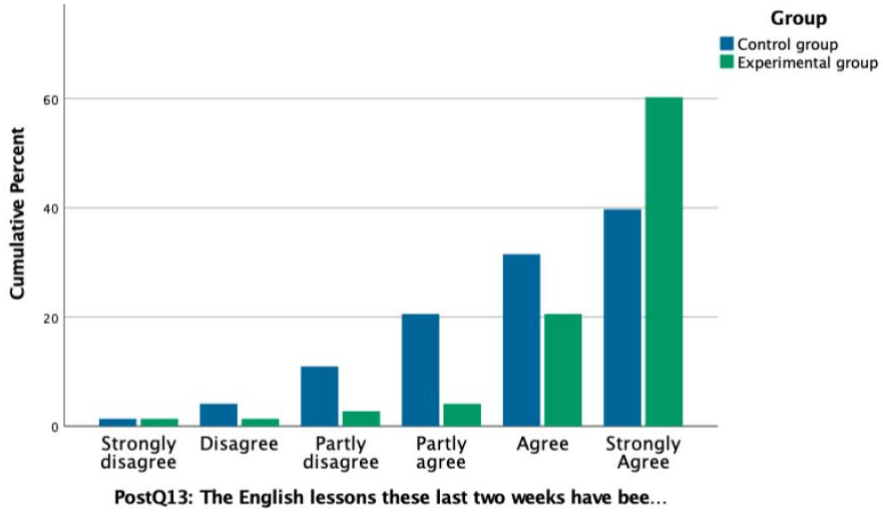
Box plot Language Attitudes Between Groups



4.2.2.3 Additional attitude statement – enjoyment

The additional question that was added to the post-test survey showed that the pupils in the experimental group had much more positive attitudes towards the lessons during the two-week treatment. The statement they answered was “the English lessons these past two weeks have been fun” (translated from Norwegian). As illustrated in Figure 9, there was a very high percentage of the pupil in the experimental group that agreed to some extent to that statement, approximately 96%, whereas around 72% of the pupils from the control group, who had regular English lessons, agreed to some degree to the statement.

Figure 9
Bar Chart L-attitude Statement



4.3 Language anxiety

This section will be divided up into two main sections according to research questions two and five. But firstly, the items making up the multi-item scale L-anxiety must be tested for internal reliability to check if they can be added together. The Cronbach's alpha test was run to test for internal reliability. It showed that the multi-item scales for L-anxiety have high internal reliability (.848). Consequently, the item scores could be added together to form the multi-item scales L-anxiety.

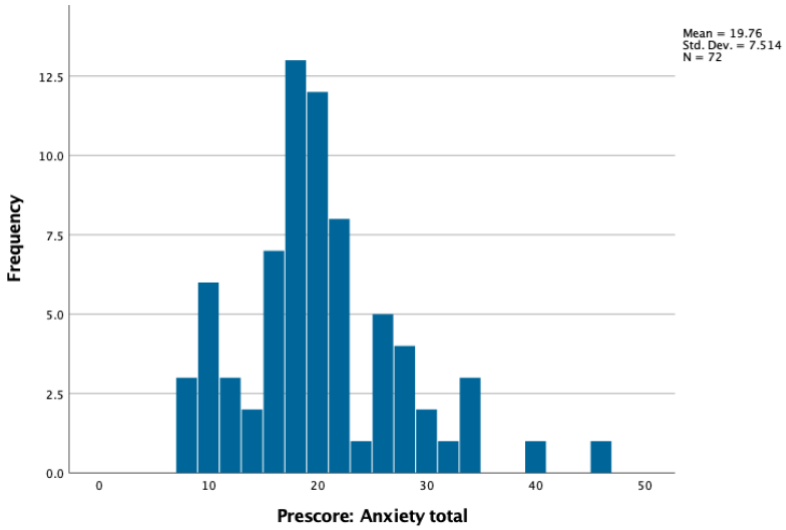
4.3.1 The language anxiety of the sample

To comment on research question two, the language anxiety scores of the sample need to be investigated. The L-anxiety score is a total score of eight items, four regarding English use anxiety and four regarding English class anxiety. The construct L-anxiety has a minimum score of eight and a maximum score of 48, where a high score indicates high levels of L-anxiety.

The participants (N = 72) average score for L-anxiety is a little below half (M = 19.76, SD = 7.514), with a range of eight to 46. This average score indicates that more pupils report having lower levels of L-anxiety than higher, however, the range shows that some pupils report having no L-anxiety, while others report having quite high levels of L-anxiety. Figure 10 illustrates this by showing that the data for L-anxiety is skewed to the left, which indicating that the majority of the pupils have low levels of L-anxiety. However, quite a few participants are seen to have reported having higher levels of L-anxiety, hence the right tail in the figure.

Figure 10

Histogram Language Anxiety



4.3.2 The effect of a collaborative output task (classroom debate) on the pupils' language anxiety

To comment on research question five, two main aspects must be looked at. Firstly, the difference between pre- and post-test scores for L-anxiety within the groups will be presented to look for potential differences within the groups. Secondly, the difference in scores between the two groups will be looked at. Before the tests can be run, the groups need to be checked for normal distribution. The results from the normal distribution test are displayed in Table 15 and it shows that the data was normally distributed for the control group prescore (p = .274) and the experimental group postscore (p = .490), while it was not normally distributed for the experimental group prescore (p = .025) and the control group postscores (p = .001), as assessed by the Shapiro-Wilk's test. Subsequently, the inferential statistics need to be obtained with the use of non-parametric test, namely, the Wilcoxon signed ranks test and Mann-Whitney U-test.

Table 15

Shapiro-Wilk's Test Language Anxiety

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
Group		Statistic	df	Sig.	Statistic	df	Sig.
1 Control group	Prescore: Anxiety total	.172	28	.032	.956	28	.274
2 Experimental group	Prescore: Anxiety total	.137	44	.037	.902	44	.001
1 Control group	PostScore:Anxiety total	.137	26	.200*	.909	26	.025
2 Experimental group	PostScore:Anxiety total	.076	40	.200*	.974	40	.490

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

4.3.2.1 Language anxiety pre- and post-test scores – Within groups

The Wilcoxon signed ranks test was conducted to determine whether the pupils' L-anxiety were statistically different from the pre-test to the post-test. From the control group (N = 24), 13 pupils reported having less L-anxiety on the post-test as opposed to the pre-test, eight reported having higher L-anxiety, and three reported having the same L-anxiety on both tests. There was a median decrease of 0.5 points from the pre-test (Mdn = 19.0) to the post-test (Mdn = 18.5), but this was not statistically significant, z = -1.137, p = .256 (Table 16).

From the experimental group (N = 40), 15 pupils reported having less L-anxiety on the post-test as opposed to the pre-test, 17 reported having more L-anxiety, and eight reported the same L-anxiety on both tests. There was a median increase of 0.5 points from pre-test (Mdn = 18.0) to post-test (Mdn = 18.5), but this was not statistically significant, z = -.583, p = .560 (Table 16). Figure 11 illustrates the participants differences in scores from the pre-test to the post-test.

Table 16

Wilcoxon Signed Ranks Test Language Anxiety

Group		PostScore: Anxiety total - Prescore: Anxiety total
1 Control group	Z	-1.137[b]
	Asymp. Sig. (2-tailed)	.256
2 Experimental group	Z	-.583[c]
	Asymp. Sig. (2-tailed)	.560

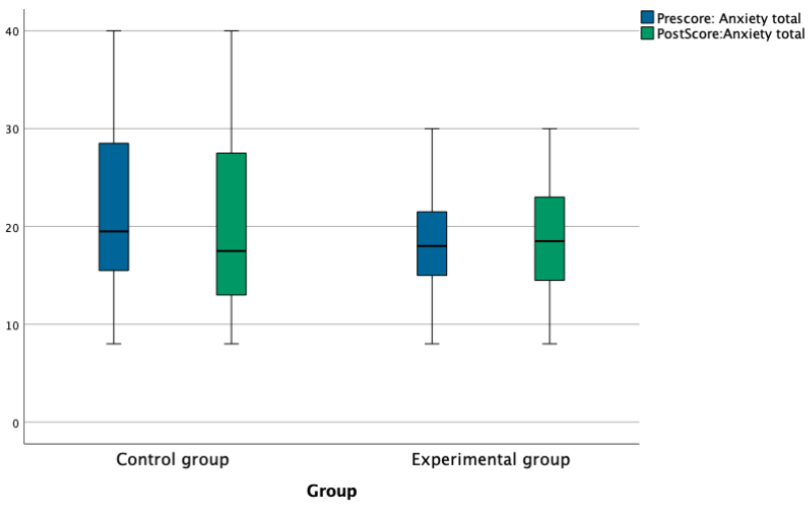
a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

c. Based on negative ranks.

Figure 11

Box plot Language Anxiety Within Groups



4.3.2.2 Language anxiety pre- and post-test scores – between groups

A Mann-Whitney U test was run to determine if there was a significant difference in L-anxiety scores between the control group and experimental group for both pre- and post-test. Distributions of the L-anxiety scores were similar, as assessed by visual inspection. For the pre-test the control group (*Mdn* = 19.0, *N* = 28,) had a higher median than the experimental group (*Mdn* = 18.0, *N* = 44), but the difference was not statistically significant, *U* = 502, *z* = -1.319, *p* = .187.

For the post-test the control group (*Mdn* = 18.5, *N* = 26) had the same median as the experimental group (*Mdn* = 18.5, *N* = 40), but the difference was not statistically significant, *U* = 496.5, *z* = -.309, *p* = .757. The results from the Mann-Whitney U Test can be seen in Table 17. Both the scores from the pre- and post-test are illustrated below in the box plots in Figure 12.

Table 17

Mann-Whitney U Test Language Anxiety

Prescore: Anxiety total across Group

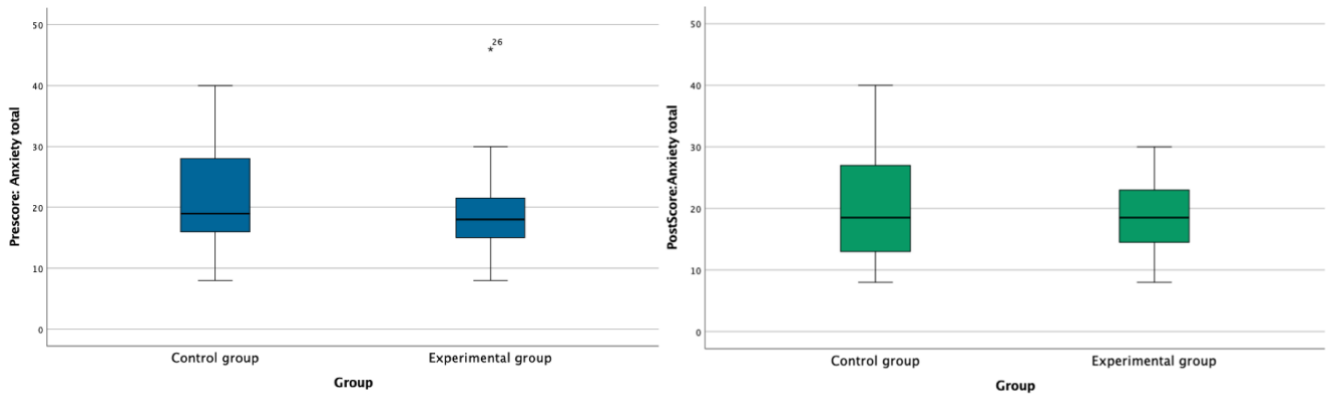
Independent-Samples Mann-Whitney U Test Summary	
Total N	72
Mann-Whitney U	502.000
Wilcoxon W	1492.000
Test Statistic	502.000
Standard Error	86.400
Standardized Test Statistic	-1.319
Asymptotic Sig.(2-sided test)	.187

PostScore:Anxiety total across Group

Independent-Samples Mann-Whitney U Test Summary	
Total N	66
Mann-Whitney U	496.500
Wilcoxon W	1316.500
Test Statistic	496.500
Standard Error	76.100
Standardized Test Statistic	-.309
Asymptotic Sig.(2-sided test)	.757

Figure 12

Box plot Language Anxiety Between Groups



4.4 The correlation between receptive vocabulary knowledge, language attitudes and language anxiety

In order to address the third research question, the pre-test for the entire sample needed to be included and tested for normal distribution through the Shapiro Wilk's test. The scores can be seen in Table 18 and they are normally distributed for vocabulary (p = .215) but not for L-attitudes (p < .001) and L-anxiety (p = .002), as assessed by Shapiro-Wilk's test. Resulting in the use of a non-parametric test to look at correlation, namely, the Spearman's rho test.

Table 18

Shapiro-Wilk's Test Overall Pre-test Scores

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pretest Vocabulary	.072	67	.200*	.976	67	.215
PreScore: Attitudes towards learning English	.166	67	.000	.866	67	.000
Prescore: Anxiety total	.131	67	.006	.937	67	.002

Table 19 shows that there is a statistical significant, negative correlation between the vocabulary pre-test score and the L-anxiety score r_s (68)= -.432, p < .001. This indicates that the participants with a high vocabulary score reported having low L-anxiety, and that the participants with a low vocabulary score reported having high L-anxiety. There is also a statistical significant, negative correlation between L-attitudes and L-anxiety r_s (71)= -.313, p = .008. This indicates that participants who had a high language attitude score, had a low L-anxiety score, and vice versa. The Spearman's rho test does, however, not show a significant relationship between language attitude and vocabulary scores r_s (68)= .146, p = .235. Figure 13 shows these correlations visually through a scatterplot. The scatterplot shows how the red dots (vocabulary and L-anxiety) and the green dots (L-attitudes and L-anxiety) have a negative correlation, whereas the blue dots (vocabulary and L-attitudes) show a slight positive correlation.

Table 19

Spearman's Rho Correlation Test

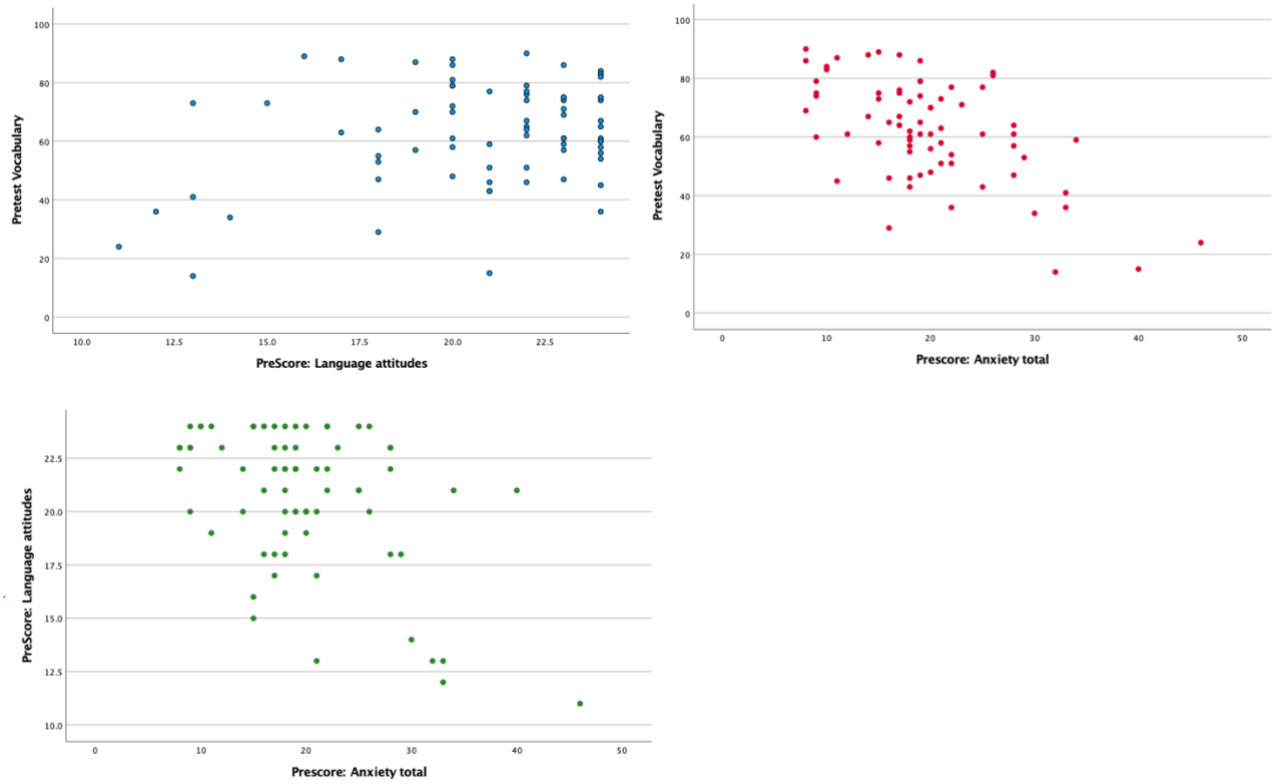
Correlations

			Pretest Vocabulary	PreScore: Attitudes towards learning English	Prescore: Language anxiety
Spearman's rho	Pretest Vocabulary	Correlation Coefficient	1.000	.146	-.432**
		Sig. (2-tailed)	.	.235	.000
		N	70	68	68
	PreScore: Attitudes towards learning English	Correlation Coefficient	.146	1.000	-.313**
		Sig. (2-tailed)	.235	.	.008
		N	68	72	71
	Prescore: Language anxiety	Correlation Coefficient	-.432**	-.313**	1.000
		Sig. (2-tailed)	.000	.008	.
		N	68	71	72

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 13

Scatterplot Correlation



5 Discussion

The main aim for this study was to investigate the pupils' vocabulary knowledge and to see how L-attitudes, L-anxiety and collaborate output tasks can affect the pupils' vocabulary knowledge. This chapter will discuss the results in light of the research questions to help reach the aims of the thesis. This chapter will be divided into five subsections, where each subsection will start with a short summary of the results relating to each research question, followed by possible explanations for the findings and connecting these findings to earlier research. Lastly, the limitations of the study and suggestions for future research will be discussed.

5.1 Norwegian 7th graders vocabulary knowledge

The first research question asks about the receptive vocabulary knowledge of Norwegian 7th graders. The results of this study indicate that the pupils' receptive vocabulary knowledge varies considerably, where the overall range stretches from 24/90 to 90/90 on the total vocabulary score. This total score provides an overview of the receptive vocabulary knowledge of the pupils on the entire test. This vast range from 24 to 90 points on the vocabulary test gives indications to the teachers that the pupils have very different vocabulary knowledge. However, the main strength of this test is that it indicates what individual the pupils should focus the vocabulary learning on, based on the scores for the levels (Webb et al., 2017, p. 34). Therefore, to further help the teachers understand how to deal with this variation in vocabulary knowledge, it is important to take a closer look at the vocabulary level scores of the pupils.

The results demonstrate that a majority of the pupils achieved full mastery of level 1.000. As mentioned in section 4.1, Webb et al. (2017) put the threshold at 29/30 point for the first three-thousand levels. This suggests that for this study, 51.4% of the pupils mastered level 1.000, whereas less than half of the pupils (17.2%) mastered level 2.000. These two levels (1.000 and 2.000) represent the high-frequency words, and overall, less than half of the pupils (16%) had full mastery over these levels. This suggests that there needs to be a greater focus on teaching high-frequency words in school. Nation (2013, pp. 9-43) stresses the importance of knowing the high-frequency words since they account for such a large portion of running words in text (90% coverage). Table 3 showed the number of words that are necessary to know in order to accomplish certain aspects of English (section 2.1). This table showed that with a mastery below 1.000 word families it is only possible to know the survival vocabulary for foreign travel and read the easiest graded readers. Learners with a mastery of levels 1.000 to 2.000 can read intermediate level graded readers and have basic speaking skills (Nation, 2013, p. 39). According to the new curriculum, the pupils should, after 7th grade, for example, be able

to listen to and understand adapted and authentic texts, read and listen to English non-fiction texts and young adult literature, and write and converse about the content (Utdanningsdirektoratet, 2020). In order to reach aims like these, the pupils need to at least have mastery of the high-frequency words, according to Nation (2013, p. 39). However, according to the findings from this study, teachers cannot assume that their 7th graders have mastery of the high-frequency words. Therefore, it is highly important that the teacher investigate what vocabulary level the learners are at and focus on getting the pupils to reach full mastery of at least the first two-thousand levels.

According to Nation (2008, pp. 1-3), the pupils will learn the high-frequency words by being involved in lessons that include all of the four strands: *Meaning-focused input, Meaning-focused output, Language-focused learning* and *Fluency development.* In the meaning-focused input strand the learners should meet vocabulary through reading and listening. This can be done by, for example, extensive reading activities or listening to stories. In this strand it is important to adjust the input to the learners' vocabulary level so that they meet both new and already known vocabulary. In the meaning-focused output strand the learners should take part in speaking and writing activities to push the boundaries of their vocabulary knowledge through activities such as problem solving and role-play. In the third strand, language-focused learning, deliberate attention is put into vocabulary and vocabulary strategies. And lastly, in the fluency development strand, the learners get more proficient at using already known vocabulary through practicing reading, listening, speaking and writing. These four strands should get equal amounts of time during the lessons; hence, each strand should get a quarter of the time to teach the learners the high-frequency words (Nation, 2008, p. 2). The first two strands, meaning-focused input and meaning-focused output, do, however, not work if there are large amounts of unknown vocabulary. The meaning-focused activities will then become more language-focused where the attention is mainly on the unknown vocabulary and not the message. Similarly, the fluency development activities are not effective if there is much unknown vocabulary (Nation, 2013, p. 3). It is, therefore, very important to know the learners' vocabulary level so that the teachers can adjust both the way to teach, and what material to use, for the different learners.

The results showed that only 5.8% of the pupils mastered level 3.000. Regardless of the scarce number of pupils who did master level 3.000, some pupils did, and since this study did not investigate levels higher than 3.000, it is not possible to tell if some pupils have mastery of an even higher level. With a mastery of the 3.000 level, the pupils should have basic listening skills, be able to read graded readers, read unsimplified texts with the help of a dictionary and watch TV (Table 2.2, Nation, 2013, p. 39). These skills come in addition to the ones they

already have achieved from mastery of the high-frequency words. These skills give sufficient coverage of the aims from the curriculum. Consequently, it would be beneficial to reach this degree of vocabulary knowledge for the 7th graders. However, very few pupils did so for this study. It is, therefore, necessary to understand how to help the learners improve their high- and mid-frequency word knowledge in order to reach these aims.

Since it is estimated that one needs to be exposed to a word eight to ten times for the learner to attain receptive knowledge of the word and the mid-frequency words only account for only around 9% of running words, these are clearly less likely to be picked up than the high-frequency words (Nation, 2013, p. 21; Schmitt & Schmitt, 2020, pp. 143-144). It is, therefore, necessary, as mentioned in section 2.1.3, to teach the mid-frequency words in a different way than the high-frequency words. According to Nation (2013, p. 27), the mid-frequency words should be taught through working with vocabulary strategies. He specifically mentions four strategies: guessing from context, using word cards, using word parts, and using a dictionary. Other researchers suggest other strategies, such as, cognitive strategies where activities can be to connect words to images or rhymes and rhythms, or making word lists or mind maps. Metacognitive strategies where the point is to reflect on the learning process through organizing and planning the learning (Gausland & Haukås, 2011). These are only a few examples of word-learning strategies. Once the learners understand how to use some of the vocabulary strategies, the teacher can provide texts with mid-frequency words and the learners can learn new vocabulary by themselves with the help of the strategies. Hence, the teachers should not teach the learners mid-frequency words, but teach them strategies to deal with unfamiliar words.

The results from this study clearly highlight the variation in vocabulary knowledge within a class. Although these results are difficult to compare to other studies due to the difference of the grades that are investigated, and the instruments used to measure, the massive range within vocabulary knowledge that is found in this study is similar to that of other studies. Olsen (2016) studied pupils in lower secondary school and found a range from 600 to 10.400 word families in his master thesis. Onyszko (2019) investigated pupils in upper secondary school and found that the minimum score was 5.100 word families, and the maximum score was 12.700 word families. Onyszko highlights the importance of acknowledging this range when teaching vocabulary. These major gaps in range, and the fact that some pupils did not master level 1.000 while others got a full score on level 3.000 in this study calls for attention. Teachers need to assess their learners to understand the variation that is within the classroom and teach vocabulary accordingly. If half of the class has full mastery of the high-frequency words, their vocabulary learning focus should be on developing vocabulary strategies and

working more independently with unfamiliar words. Whereas the teacher needs to keep working actively with the four strands with the pupils who have not yet mastered the high-frequency words. Therefore, attention must be given to the vocabulary level of the learners so that they can be taught accordingly and keep improving their vocabulary knowledge through their education.

5.2 Norwegian 7th graders language attitudes and anxiety

The second research question addresses the L-attitudes and L-anxiety of Norwegian 7th graders. The results show that a majority of the pupils reported having positive attitudes towards learning English (*M* = 20.69, *SD* = 3.283). An average score of 20.69 out of 24 is quite high. That indicates that most pupils participating in this study reported having very positive attitudes towards learning English. Yet, the scores ranged from 11 to 24, indicating that some pupils reported rather negative attitudes towards learning. This is similar to the vocabulary scores, that the range within the groups is quite big.

L-anxiety, on the other hand, shows that the average score is a little below half (*M* = 19.76, *SD* = 7.514). The maximum score for L-anxiety is 48, hence, an average score of 19.76 is close to the middle. The range stretches from 8-46, which implies that some pupils reported having no L-anxiety, since a score of 8 is the minimal score one can get that indicates having the least L-anxiety. Meanwhile, some pupils reported having almost the maximum score, which indicates high levels of L-anxiety.

These scores indicate that there are big variations in L-attitude and L-anxiety within this sample. Although teachers do not necessarily want or need to test their pupils' L-attitudes and L-anxiety, it has been seen to be important for teachers to be aware of this variance because the pupils' L-attitudes and L-anxiety affect all aspects of L2 learning. As seen in section 2.2 and 2.3, L2 motivation, hence L-attitudes, is an important aspect of both the initial appraisal of L2 learning, as well as the driving force to sustain the learning process (Dörnyei, 1998, p. 117; Tseng & Schmitt, 2008). Similarly, previous research indicates that L-anxiety has a reciprocal relationship with language achievements (Gardner, 2010). L-attitudes and L-anxiety are, therefore, important beyond their own construct, subsequently, it can be helpful to know of ways to make the pupils feel more positive towards learning English and less anxious, in order to improve their language achievements.

Section 2.4.4 presents previous studies where it has been found that project and collaborative work can be highly motivating activities for the pupils and that these activities

can get the entire class caught up in a tide of motivational energy (Henry et al., 2018). Similarly, it presents research on how oral activities can feel less anxiety-provoking for the pupils with practice, group work and a warm and friendly environment (Byrne et al., 2012; Hashemi, 2011; Ozturk & Gurbuz, 2014; Young, 1990). These are elements that can be useful in a classroom situation if the pupils either have little motivation or high L-anxiety. There are, of course, many other aspects that can help the pupils, and teachers most likely have their own thoughts of what one can do in their classroom to work with issues regarding L-attitudes and L-anxiety.

5.3 Vocabulary knowledge and language attitudes and anxiety

The third research question addresses the correlation between receptive vocabulary knowledge, L-attitudes and L-anxiety. The results show a statistically significant, negative correlation between receptive vocabulary knowledge and L-anxiety (r_s (68) = -.432, p < .001). This suggests that the pupils who had the most extensive vocabulary knowledge also reported having low levels of L-anxiety, and vice versa. This correlation supports the reciprocal relationship between language achievement and L-anxiety in Gardner's socio-educational model (2010, p. 88). This suggests that the vocabulary knowledge and L-anxiety go hand in hand, where one influences the other. Hence, if a pupil knows very little vocabulary, it is likely that their L-anxiety increases, and if they have high levels of L-anxiety, they will struggle to obtain an extensive vocabulary.

This finding is also in line with previous studies on the connection between vocabulary learning and L-anxiety. They showed that L2 anxiety can influence both vocabulary learning and production and that L2 anxiety can lead to poor performances on vocabulary tests, as well as lead to a longer time to learn new vocabulary and a longer time completing vocabulary test (MacIntyre, 2017; MacIntyre & Gardner, 1989). High levels of anxiety can often come across as fear of using the language, and as discussed in section 2.2, vocabulary needs to be learned both implicitly and explicitly, through input and output. Consequently, if a pupil is afraid to engage with the language, and, therefore, does not get involved with the language more than they have to, that can greatly affect the pupils' ability to learn new vocabulary. This can either be a positive circle with low L-anxiety and high vocabulary knowledge, or a negative circle with high L-anxiety and low vocabulary knowledge. However, there are ways to affect this circle by focusing on improving their vocabulary knowledge and L-anxiety through the methods discussed in section 5.1 and 5.2. This significant correlation found for this sample, highlights the importance of vocabulary knowledge and L-anxiety and their interrelatedness.

A second significant, negative correlation, was found between L-attitudes and L-anxiety (r_s (71)= -.313, p = .008). This is in line with the model presented in chapter 2 (Figure 2), where L-attitudes and L-anxiety have been tied together as representing the broader view of language motivation. In Tseng and Schmitt's model (2008), both anxiety and attitudes make up two of the three indicators for motivation. However, in Gardner's socio-educational model there is no direct relationship between L-anxiety and any of the parts of motivation, but the relationship is tied together with language achievements in the middle. The findings from this thesis suggest a correlation between L-attitudes and L-anxiety. This can indicate that if one works to improve the L-attitudes, the L-anxiety can be improved as a byproduct, and vice versa. Consequently, it strengthens the notion that working with these aspects are important not only to that specific aspect, but that it also can affect other aspects of learning.

Lastly, no statistically significant correlation was found between receptive vocabulary knowledge and L-attitudes (r_s (68) = .146, p = .235). In a somewhat related study, Jakobsson (2018), found that the pupils' English grades correlated with the pupils' L-attitudes. This, however, might be due to the broad specter of English grades versus English vocabulary, or to the different measuring tools used to measure L-attitudes. Since this study only investigated one out of three aspects of motivation, it was less likely to find a strong correlation with vocabulary, because, as Gardner (2010) states, the individual components are not sufficient to represent motivation, but it can express the pleasure and enjoyment associated with L2 learning.

5.4 Collaborative output task and receptive vocabulary knowledge?

The fourth research question asks about the effect collaborative output tasks have on receptive vocabulary knowledge. To begin with, the results show that both groups had significant improvements from the pre-test to the post-test ($p < .05$). This increase in scores can indicate several different things, such as a practice effect, where the pupils simply perform better due to gaining more experience with taking the test or a test effect because the pupils remembered items from the test and checked it when they got home. However, as discussed in section 3.8, this was considered unlikely. Considering that the pupils were given a thorough explanation of the test and assistance while taking the test, the practice effect should be limited. Similarly, the test effect should be limited due to the large number of words and sentences on the test, hence, posing a rather small chance of the pupils remembering several words to search for at home. The pupils could also have had a better day during one of the tests, although several steps were taken to reduce boredom and fatigue so that the scores would reflect their true knowledge and not a good or bad day, it is not possible to eliminate that threat entirely when investigating

humans. Therefore, that could be a potential factor as to why they got an increase in scores. The pupils might also have gotten a better score due to them actually learning more words during the two weeks. Regardless of the steps taken to reduce confounding variables, it is not possible to completely rule out other variables that might have affected these outcomes, and we can, therefore, not be sure that these improvements are due to an actual improvement in vocabulary knowledge.

The results show an average increase of 2.7 points for the control group, and 2.4 points for the experimental group. This indicates that the control group in fact had a larger increase than the experimental group. Nevertheless, the results also show that there were no statistically significant differences between the control and experimental group on either test ($p > .05$). This suggests two things: firstly, there were no pre-existing differences between the groups on the vocabulary level test, and secondly, there were no significant differences for the groups on the post-test. Hence, even though there was a slight difference in the groups increase from pre-test to post-test, it was not large enough to be statistically significant for either test. This suggests that for this study, the collaborative output task did not elicit a greater increase in vocabulary knowledge than the regular English lessons. However, the collaborative output task did not elicit a significantly lower increase than the regular English lessons either. As mentioned, other factors might have influenced the results and it is, therefore, not possible to be sure if the treatment had anything to do with the increase in scores.

These findings are different from previous research discussed in section 2.4, where the pupils engaging in collaborative output tasks outperform the control groups in vocabulary acquisition and retention (Kim, 2008; Nassaji & Tian, 2014; Sun, 2017). A reason for these differences in findings may be that these previous studies have used target vocabulary that the groups are actively working with during the study. In their studies, it was more likely that the learners would pick up new vocabulary, since they worked with the target vocabulary that they were tested on. In this current study, it was not certain that the pupils would get an increase in vocabulary scores since they did not work specifically with the vocabulary that was tested. Hence, the test used in this study might have been too general to pick up such changes.

Regardless, the results from this study show a significant increase in scores for both groups, what we cannot be sure of, is why these scores increased. However, previous research has shown that collaborative output tasks, such as debate, can improve the learners' vocabulary knowledge. Additionally, a classroom debate project involves at least two of the four strands that Nation (2008, pp. 1-3) suggest need to be present to teach and learn the high-frequency words: *meaning-focused input* and *meaning-focused output.* In a debate, the pupils work with

English material both by reading and listening to articles, and by taking part in writing and speaking activities, both within their groups, and in front of the class.

5.5 Collaborative output task and language attitudes and anxiety

The fifth research question addresses whether the collaborative output task had an effect on the L-attitudes and L-anxiety of the pupils. The results show that the control group got a median decrease of 0.5 points in positive attitudes towards learning English from the pre-test to the post-test, while the experimental group got a median increase of 1 point. Yet, none of the results were statistically significant ($p > .05$). This slight increase in L-attitudes seen in the experimental group is similar to findings from previous research which has found that working collaboratively to produce output over a longer period of time can be highly motivating for the learners (Henry et al., 2018). However, people have very different interests that make them feel pleasure and enjoyment, and therefore, debate might not be an activity that makes every pupil feel more enjoyment in the L2 learning. Consequently, it is important to have varied lessons so that every pupil can feel an increase in attitudes towards learning English occasionally.

No significant difference was found between the control and experimental group on the L-attitude pre-test or post-test ($p > .05$). Even though previous studies have shown that when individuals work together in groups, their motivational levels can significantly exceed what they would have achieved if they had worked independently (Dörnyei & Ushioda, 2011), that was not picked up on in this study. There are several reasons for why the results might not have shown a significant difference between the two groups, and one reason might be that two weeks of debate is not enough to influence the general attitudes of L2 learning. Although this motivational increase was not picked up on through the survey, it was commented on throughout the study by teachers, principals and parents. Comments such as "I have never seen my child so engaged in something before" and "the pupils are so excited to show you what they have found out" suggests a different view of motivation that did not come through in the survey. One of the teachers from the study contacted me and said that during the student-teacher conference, months after the project, many of the pupils had mentioned the debate project as something they wish they could have more of, because it was such a fun way to learn English. Additionally, 96% (42/44) of the pupils from the experimental group agreed that the lessons during the two-week treatment period had been fun, whereas only 72% (21/29) of the pupils from the control group agreed to the same statement. This indicates that the debate project was enjoyable and motivating for the pupils, but it was not enough to significantly influence their overall attitudes towards learning English.

When it comes to L-anxiety, the control group had a median decrease of 0.5 points while the experimental group had a median increase of 0.5 points, none of the results were statistically significant. This, however, indicates that the experimental group reported having slightly more L-anxiety, whereas the control group reported slightly less L-anxiety after the two weeks. This finding is in line with several of the previous studies on L-anxiety and oral tasks (Horwitz et al., 1986; Young, 1990). Given that speaking in a foreign language is considered the most anxiety-provoking activity, it is not very surprising that the experimental group reported having slightly more L-anxiety after the treatment. Nevertheless, no significant difference was found between the control group and experimental group on either the pre-test or the post-test ($p >$ .05). There might be several reasons for why no significant difference was found, and the main one might be the same as for the L-attitudes, that the two-week debate treatment was not enough to significantly influence general L-anxiety, at least not enough to be picked up on a survey like the one implemented in this study. It is, therefore, not possible to be entirely sure if a collaborative output task, such as a classroom debate, can affect the pupils' L-anxiety, and in what way it will be affected, from the results of this study. As mentioned in section 2.4 and 2.5, previous research indicates that on one side, collaborative output tasks can be very anxiety-provoking, while on the other side, it can be less anxiety-provoking if certain elements are present, such as time to prepare and being in a warm and friendly environment (Byrne et al., 2012; Hashemi, 2011; Ozturk & Gurbuz, 2014; Young, 1990). Since practice is mentioned as one of the key elements to making the speaking process less anxiety-provoking, it is important that the teachers keep putting the pupils in situations where they can practice speaking where they feel safe and get a feeling of mastery. Especially since this study has indications of significant correlations between L-anxiety and both vocabulary and L-attitudes.

5.6 Limitations

The limitations of this study are divided into two sections: limitations of overall design of the entire study and of the quasi-experimental design. This section will present possible limitations to this study.

5.6.1 Limitations of overall design

The overall study and design has some limitations that will be presented in this section. The first limitation is the nonrandom sample, which is, as explained in section 3.8, difficult to achieve in L2 research. Since one of the aims for this study was to find results that could be

useful for teachers in their classroom it was considered necessary to keep the classes intact. This could, however, have influenced the results since the pupils participating in this study might have been better or worse at vocabulary than the general Norwegian 7th grader or had much higher L-anxiety. This is a limitation that is present in order to keep away other limitations, such as an unrealistic classroom setting.

The other limitation was the size of the test and survey. The vocabulary test measured the pupils' receptive vocabulary knowledge on the basis of the first three-thousand frequency levels. This choice was made due to the vocabulary knowledge Norwegian 7th graders need. Yet, it does not adequately get the entire image of the pupils' vocabulary knowledge, firstly because it did not include levels above 3.000, and secondly because it only measured vocabulary level and not size. The reason for not including more tests and higher levels were threats to validity due to fatigue and boredom, as discussed in section 3.8. Regardless, this does limit the scope of receptive vocabulary knowledge.

5.6.2 Limitations of quasi-experimental design

Along with the limitations to the overall design, the quasi-experimental part of this study has some additional limitations. Firstly, the design would have benefitted from a more structured plan for the control group. Since no one monitored the control groups there was no telling if they had oral activities or vocabulary focus learning, even though they were told to stay away from activities of that sort. It also meant that all four control groups could have had completely different lessons. However, since much of the vocabulary learning, especially of the high-frequency words, happen incidentally, all engagement with English could potentially be a vocabulary learning situation. Consequently, the only way to ensure that pupils do not take part in activities that could lead to vocabulary learning during the two weeks, would have been if they did not have English lessons for two weeks nor engage with English outside of school. This is very difficult to achieve, as well as ethically problematic. Hence, it was not considered an option in this study.

Secondly, the issue with other variables affecting the scores was not just present with the control group, but also with the experimental group. In a quasi-experimental study, the participants will be influenced by other sources through the project. Confounding variables such as boredom, tiredness and involvement in English activities outside of school can influence the results without being picked up. It is impossible to completely eliminate these variables. Nonetheless, steps can be taken to limit them. Some aspects were implemented in this study to limit this threat, such as, taking the test around the same time of the day both times

and having shorter tests. Still, this threat would have been further limited if the pupils took the test during the same day both times. This was not possible during this study due to scheduling issues.

Lastly, other limitations had to do with the tests/surveys that were used. One issue was that the vocabulary level test used in this thesis investigated general vocabulary knowledge, which suggests that the pupils did not encounter this vocabulary during the two-week lessons. This was not an issue of the investigation of vocabulary knowledge of 7th graders in general, since the point of that part of the thesis was to investigate the pupils' general vocabulary. However, it was more of an issue for the treatment section of the study, since the point then was to look for effects of the treatment. This study did not investigate whether collaborative output tasks were a better way to pick up specific vocabulary than regular English lessons, but it investigated whether the pupils got access to a larger understanding of general vocabulary. With the implementation of the level test, the pupils' overall general vocabulary was tested, and there were no real indications from the previous research that suggested that the pupils would pick up general vocabulary from two weeks of collaborative output tasks. As mentioned, several times before, the decision to measure general vocabulary was based on many factors, such as time issues, logistical issues and the importance of vocabulary frequency, but for the quasi-experimental aspect of this study, it would potentially have been more beneficial to investigate targeted vocabulary.

5.6.3 Further research

Given the limitations of this study, and the limited number of similar studies conducted in the Norwegian elementary setting, I highly urge further research to be conducted both to investigate the general receptive vocabulary knowledge, L-attitudes and L-anxiety of Norwegian 7th graders. As well as to look into how collaborative output tasks can affect these aspects. This study has touched upon several areas that have been under-researched in the Norwegian context, and especially with elementary grade pupils.

Considering the nonrandom sample and intact groups used in this study, further research into the different aspects would help to further generalize the results. According to the results from this study, there seems to be a big gap between the vocabulary knowledge Norwegian 7th graders should have, and what they actually have. Given the great importance of knowing the high-frequency word, and the variance in frequency knowledge that has been revealed in this thesis, I especially urge future researchers to keep investigating the vocabulary level of Norwegian pupils at different grades.

Additionally, it would be interesting to investigate to what extent targeted work with the four strands over a time period would affect pupils' knowledge of high-frequency words. As well as how a focus on vocabulary strategies can affect the knowledge of mid-frequency words for older pupils.

Lastly, it would have been interesting to further investigate the connection between vocabulary knowledge, L-attitudes, L-anxiety and characteristics of the pupils, since that analysis could not be made within the scope of this study. It would be interesting to see how the pupils' background characteristics and their engagement with English outside of school connects with their vocabulary knowledge, L-attitudes and L-anxiety in the Norwegian context.

6 Conclusion

This master thesis set out to investigate the receptive vocabulary knowledge of Norwegian 7th graders. This was done both through investigating the vocabulary knowledge by itself, and by looking at how aspects like L-attitudes, L-anxiety and a collaborative output task affected the pupils and their vocabulary knowledge. By administering pre-tests to 74 7th graders, this thesis was able to explore the pupils' initial vocabulary knowledge, L-attitudes and L-anxiety. Further, by dividing the pupils into two groups: control- and experimental groups, and administering a treatment and a post-test, this study investigated how the pupils' vocabulary knowledge, L-attitudes and L-anxiety was affected by a collaborative output task.

The first aspect was to investigate the 7th graders vocabulary knowledge. The results suggested that the pupils' receptive vocabulary knowledge varies considerably within the sample, where some pupils did not master the high-frequency levels, while others got a full score on level 3.000 (mid-frequency). These results only provide information regarding the sample; however, similar findings were found in other studies, suggesting that this might have indications beyond this sample. This vast variance within the vocabulary levels of the pupils has implications for English teachers in Norway. Not only did the results suggest a variation within the sample, but a huge gap was found between the what the pupils should know according to the curriculum and what they do know. It indicates that the teachers need to investigate their learners' vocabulary level in order to target the vocabulary teaching and the materials to the different frequency levels through the four strands and vocabulary strategies. In that way, the pupils can keep increasing their vocabulary knowledge through their education and reach the aims from the curriculum.

The results for the pupils' L-attitudes showed that the pupils reported positive L-attitudes. Nevertheless, the variation within the sample was yet again vast, ranging from 11 to 24. The L-anxiety scores showed that the pupils reported somewhat low levels of L-anxiety, but the variation was large, ranging from 8-46. Both L-attitudes and L-anxiety are aspects that can affect the overall L2 learning. L-attitudes are important both for the initial incentive to learn L2, as well as the driving force to keep learning. L-anxiety has a reciprocal relationship with language achievements, where high levels of L-anxiety can negatively affect the learning process and low levels can positively influence the learning process. Considering the big variation of scores in the sample and the importance of L-attitudes and L-anxiety in the learning process, it is important that teachers are aware of how it can influence the language learning, and how to deal with it if they encounter low levels of L-attitudes or high levels of L-anxiety in their classroom.

The second aspect was to investigate how a collaborative output task, a classroom debate, affected the pupils' receptive vocabulary knowledge, L-attitudes and L-anxiety. These results showed significant improvement on the vocabulary test for both groups from the pre-test to the post-test, but no significant difference between the control and experimental group. This indicates that the pupils from both groups got an increase in vocabulary scores from two weeks of English lessons. However, it does not tell us if the increase was specifically due to the lessons or other variables. Considering the previous research that has been discussed, and the different ways of learning new vocabulary, it seems likely to assume that implementing classroom debate projects, or other collaborative output tasks, will help the learners increase their vocabulary knowledge, even though the results from this study did not explicitly indicate that.

The results for the L-attitudes and L-anxiety showed no significant differences within the groups, or between the groups for either of the surveys. However, the additional statement that the pupils answered, whether the English lessons during the last two weeks had been fun, indicated that the pupils found the debate project enjoyable and motivating. This indicates that even though a two-week debate project might not elicit a significant improvement in overall L-attitudes, it can affect the pupils' L-attitudes in that moment.

Regardless of the limitations of this study, it aimed to address a gap in several areas of L2 research in the Norwegian elementary grade setting. It set out to investigate the pupils' vocabulary knowledge, and to see how L-attitudes, L-anxiety and a collaborative output task might influence that. This thesis might not have been able to fill the research gap, but it has made a contribution with the vast variation in scores that was found in the sample. This finding further highlights the importance to keep investigating the pupils' vocabulary knowledge in order to gain awareness on how to adjust the learning accordingly and to keep filling the research gap.

References

Bachman, L. F. (2004). *Statistical Analyses for Language Assessment Book*: Cambridge University Press.

Bjørke, C. (2018). Vocabulary learning. In H. Bøhn, M. Dypedahl, & G.-A. Myklevold (Eds.), *Teaching and learning English*. Oslo: Cappelen Damm akademisk.

Brown, J. D. (2011). Likert items and scales of measurement? *SHIKEN: JALT Testing & Evaluation SIG Newsletter, 15*(1), 10-14. doi: 10.1037/h0081614

Byrne, M., Flood, B., & Shanahan, D. (2012). A Qualitative Exploration of Oral Communication Apprehension. *Accounting education (London, England), 21*(6), 565-581. doi:10.1080/09639284.2012.725636

Clément, R., Gardner, R. C., & Smythe, P. C. (1977). Motivational variables in second language acquisition: A study of francophones learning English. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 9*(2), 123.

Cobb, T. (2007). Computing the Vocabulary Demands of L2 Reading. *Language learning & technology, 11*(3), 38-64. doi:10125/44117

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education*. In L. Manion & K. R. B. Morrison (Eds.), (8th ed.).

Converse, J. M., & Presser, S. (1986). *Survey questions:handcrafting the standardized questionnaire*(Vol. 63).

Council of Europe. Common European Framework of Reference for Languages: learning, teaching, assessment. Retrieved from https://www.eui.eu/documents/servicesadmin/languagecentre/cef.pdf

Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge: Cambridge University Press.

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2020). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types of input are important? *Bilingualism (Cambridge, England), 23*(1), 171-185. doi:10.1017/S1366728918001062

De Wilde, V., & Eyckmans, J. (2017). Game on! Young learners' incidental language learning of English prior to instruction. *Studies in second language learning and teaching, 7*(4), 673-694. doi:10.14746/ssllt.2017.7.4.6

Donato, R. (1994). Collective scaffolding in second language learning. In J. P. Lantolf & G. Appel (Eds.), *Vygotskian approaches to second language research* (pp. 33-56): Ablex publishing corporation.

Dörnyei, Z. (1994). Motivation and Motivating in the Foreign Language Classroom. *The Modern language journal (Boulder, Colorado), 78*(3), 273. doi:10.2307/330107

Dörnyei, Z. (1998). Motivation in second and foreign language learning. *Language teaching, 31*(3), 117-135. doi:10.1017/S026144480001315X

Dörnyei, Z. (2001). *Motivational Strategies in the Language Classroom*. doi:10.1017/CBO9780511667343

Dörnyei, Z. (2007). *Research methods in applied linguistics : quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.

Dörnyei, Z. (2010). *Questionnaires in second language research : construction, administration, and processing* (2nd ed.). New York: Routledge.
Dörnyei, Z. (2020). From Integrative Motivation to Directed Motivational Currents: The Evolution of the Understanding of L2 Motivation over Three Decades. In *The Palgrave Handbook of Motivation for Language Learning* (pp. 39-69) Cham: Springer International Publishing.
Dörnyei, Z., & Ryan, S. (2015). *The Psychology of the Language Learner Revisited*. London: London: Taylor and Francis.
Dörnyei, Z., & Ushioda, E. (2011). *Teaching and Researching: Motivation*: Routledge.
Gardner, R. C. (1985). The Attitude/Motivation Test Battery: Technical Report. University of Western Ontario.
Gardner, R. C. (2004). Attitude/Motivation Test Battery: International AMTB Research Project. The University of Western Ontario.
Gardner, R. C. (2010). *Motivation and second language acquisition: The socio-educational model* (Vol. 10): Peter Lang
Gardner, R. C. (2019). The Socio-educational Model of Second Language Acquisition. In *The Palgrave Handbook of Motivation for Language Learning* (pp. 21-37). Cham: Springer International Publishing
Gardner, R. C., & Lambert, W. E. (1972). Attitudes and Motivation in Second-Language Learning (pp. 63-84). Psychology Press
Gausland, A., & Haukås, Å. (2011). Bruk av ordlæringsstrategier blant norske tyskelever. *Acta didactica Norge, 5*(1). doi:10.5617/adno.1064
Gjerde, E. (2020). *Language anxiety in oral activities in Norwegian lower secondary EFL classroom: Teachers' and their students' beliefs, practices and experiences.* University of Stavanger, Norway,
Hashemi, M. (2011). Language stress and anxiety among the English language learners. *Procedia-Social and Behavioral Sciences, 30*, 1811-1816. doi:10.1016/j.sbspro.2011.10.349
Helness, H. L. (2012). A comparative study of the vocabulary of 7th and 10th graders in scripts from the National Tests of Writing in English. In A. Hasselgreen, I. Drew, & B. Sørheim (Eds.), *The Young Language Learner: Research-based Insight into Teaching and Learning* (pp. 145-158).
Henry, A., Korp, H., Sundqvist, P., & Thorsen, C. (2018). Motivational Strategies and the Reframing of English: Activity Design and Challenges for Teachers in Contexts of Extensive Extramural Encounters. *TESOL quarterly, 52*(2), 247-273. doi:10.1002/tesq.394
Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern language journal, 70*(2), 125-132. doi:10.2307/327317
Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition. *Language learning, 51*(3), 539-558. doi:10.1111/0023-8333.00164
Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on psychological science, 2*(3), 260-280. doi:10.1111/j.1745-6916.2007.00044.x

Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in second language acquisition*, 541-577. doi:10.1017/S0272263102004023

Jakobsson, J. (2018). *A study of the types, frequency and perceived benefits of extramural activities on Norwegian 10th graders' development of English as a foreign language.* University of Stavanger, Norway,

Johns, R. (2005). One size doesn't fit all: Selecting response scales for attitude items. *Journal of Elections, Public Opinion & Parties, 15*(2), 237-264. doi:10.1080/13689880500178849

Kim, Y. (2008). The Contribution of Collaborative and Individual Tasks to the Acquisition of L2 Vocabulary. *The Modern language journal (Boulder, Colo.), 92*(1), 114-130. doi:10.1111/j.1540-4781.2008.00690.x

Knight, S. (1994). Dictionary Use While Reading: The Effects on Comprehension and Vocabulary Acquisition for Students of Different Verbal Abilities. *The Modern language journal (Boulder, Colorado), 78*(3), 285-299. doi:10.1111/j.1540-4781.1994.tb02043.x

Krashen, S. D. (1981). *Second language acquisition and second language learning*.

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 learnign: Research-based insights into teaching and learning* (pp. 131-143).

Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R*(2nd ed.). London: Routledge.

Lefever, S. C. (2010). English skills of young learners in Iceland:"I started talking English when I was 4 years old. It just bang… just fall into me.".

Macaro, E. (2003). *Teaching and learning a second language : a review of recent research*. London: Continuum.

MacIntyre, P. D. (2017). An Overview of Language Anxiety Research and Trends in its Development. In C. Gkonou, Daubney, M., & Dewaele, J. M. (Ed.), *New insights into language anxiety* (pp. 11-30): Multilingual Matters.

MacIntyre, P. D., & Gardner, R. C. (1989). Anxiety and second-language learning: Toward a theoretical clarification. *Language learning, 39*(2), 251-275. doi:10.1111/j.1467-1770.1989.tb00423.x

MacIntyre, P. D., Ross, J., & Clément, R. (2019). Emotions are motivating. In M. Lamb, K. Csizér, A. Henry, & S. Ryan (Eds.), *The Palgrave Handbook of Motivation for Language Learning* (pp. 183-202): Springer.

Mackey, A., & Gass, S. M. (2015). *Second Language Research: Methodology and Design*. London: Routledge.

Myhre, T. S., & Fiskum, T. A. (2020). Norwegian teenagers' experiences of developing second language fluency in an outdoor context. *Journal of adventure education and outdoor learning, ahead-of-print*(ahead-of-print), 1-16. doi:10.1080/14729679.2020.1769695

Nassaji, H., & Fotos, S. S. (2011). Focus on Grammar through Collaborative Output Tasks. In *Teaching Grammar in Second Language Classrooms : Integrating Form-Focused Instruction in Communicative Context* (pp. 113-128): Routledge.

Nassaji, H., & Tian, J. (2014). The Role of Language Coproduction in Learning English Vocabulary. *Procedia, social and behavioral sciences, 143*, 794-798. doi:10.1016/j.sbspro.2014.07.478

Nation, I. S. P. (2008). *Teaching vocabulary : strategies and techniques*. Boston, Mass: Heinle.

Nation, I. S. P. (2013). *Learning Vocabulary in Another Language*. Cambridge University Press.

Norušis, M. J. (2008). *SPSS statistics 17.0 : guide to data analysis*. Upper Saddle River, N.J: Prentice Hall.

Olsen, C. L. (2016). Keywords and Individual Differences – in relation to vocabulary acquisition and text understanding. In *Nøkkelord og Individuelle forskjeller – sett opp imot vokabularlæring og tekstforståelse*: Bergen University College.

Onyszko, M. (2019). *Learners' English Vocabulary Size in Upper Secondary School. A study of receptive and productive vocabulary size in Vg1 in Norway.* University of Stavanger, Norway,

Ozturk, G., & Gurbuz, N. (2014). Speaking anxiety among Turkish EFL learners: The case at a state university. *The journal of language and linguistic studies, 10*(1), 1-17.

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language, 22*(1), 31–55

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

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language teaching research, 12*(3), 329-363. doi:10.1177/1362168808089921

Schmitt, N. (2010). *Researching vocabulary: a vocabulary research manual*. Basingstoke: Palgrave Macmillan.

Schmitt, N. (2014). Size and Depth of Vocabulary Knowledge: What the Research Shows: Size and Depth of Vocabulary Knowledge. *Language learning, 64*(4), 913-951. doi:10.1111/lang.12077

Schmitt, N., & Schmitt, D. (2020). *Vocabulary in language teaching*: Cambridge university press.

Schnurer, M., & Snider, A. (2001). *Many Sides: Debate Across the Curriculum*: New York: International Debate Education Association.

Semerik, R. (2016). *Effects of using songs in EFL classes on levels of engagement and anxiety of upper elementary school students.* Tartu Ülikool.

Sun, C.-H. (2017). The value of picture-book reading-based collaborative output activities for vocabulary retention. *Language teaching research: LTR, 21*(1), 96-117. doi:10.1177/1362168816655364

Sundqvist, P. (2009). *Extramural English Matters: Out-of-School English and Its Impact on Swedish Ninth Graders' Oral Proficiency and Vocabulary.* (Dissertation). Karlstad University Studies, Retrieved from http://www.diva-portal.org/smash/get/diva2:275141/FULLTEXT03.pdf

Sundqvist, P., & Wikström, P. (2015). Out-of-school digital gameplay and in-school L2 English vocabulary outcomes. *System (Linköping), 51*, 65-76. doi:10.1016/j.system.2015.04.001

Swain, M. (1985). Communicative competence some roles of comprehensible input and output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in Second Language Acquisition*: Newbury House Publishers.

Swain, M. (1995). Three functions of output in second language learning. In H. G. C. Widdowson, G Seidlhofer, B (Ed.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.

Swain, M. (2005). The output hypothesis: theory and research. In H. Eli (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 471-484): Taylor & Francis Group.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied linguistics, 16*(3), 371-391. doi:10.1093/applin/16.3.371

Tseng, W.-T., & Schmitt, N. (2008). Toward a Model of Motivated Vocabulary Learning: A Structural Equation Modeling Approach. *Language learning, 58*(2), 357-400. doi:10.1111/j.1467-9922.2008.00444.x

Utdanningsdirektoratet. (2020). *Læreplan i engelsk*. (ENG01-04). Utdanningsdirektoratet Retrieved from https://data.udir.no/kl06/v201906/laereplaner-lk20/ENG01-04.pdf?lang=nob

Vilkaitė-Lozdienė, L., & Schmitt, N. (2020). Frequency as a Guide for Vocabulary Usefulness: High-, Mid-, and Low-Frequency Words. In *The Routledge handbook of vocabulary studies* (1 ed., pp. 81-96): Routledge.

Vygotsky, L. S. (1978). *Mind in society : the development of higher psychological processes*. Cambridge, Mass: Harvard University Press.

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL-International Journal of Applied Linguistics, 168*(1), 33-69. doi:10.1075/itl.168.1.02web

Victoria University of Wellington, Vocabulary tests. Retrieved from https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-tests

Wilkins, D. A. (1972). *Linguistics in language teaching*: E. Arnold, 1973.

Young, D. J. (1990). An investigation of students' perspectives on anxiety and speaking. *Foreign Language Annals, 23*(6), 539-553.

Appendices

Appendix 1: Questionnaire in Norwegian

Spørreskjema engelsk

1. Kjønn

Markér bare én oval.

- ◯ Gutt
- ◯ Jente
- ◯ Vil ikke svare

2. Hvor ofte snakker du disse språkene hjemme?

Markér bare én oval per rad

	Aldri	Nesten aldri	Sjeldent	Ofte	Nesten alltid	Alltid
Norsk	◯	◯	◯	◯	◯	◯
Engelsk	◯	◯	◯	◯	◯	◯
Annet	◯	◯	◯	◯	◯	◯

3. Prater du vanligvis engelsk med en i slekten din eller en venn, og i så fall, hvor ofte?

Markér bare én oval.

- ◯ Ja, daglig
- ◯ Ja, noen ganger i uken
- ◯ Ja, noen ganger i måneden
- ◯ Ja, noen ganger i året
- ◯ Nei

4. Hvor mange timer bruker du på de ulike tingene hver dag?

Markér bare én oval per rad

	Mer enn 4 timer	Mellom 2-4 timer	Mellom 1-2 timer	Under 1 time	Aldri
Engelsk serie/film	◯	◯	◯	◯	◯
Engelsk spill	◯	◯	◯	◯	◯
Engelsk på Youtube	◯	◯	◯	◯	◯
Engelsk sosiale medier	◯	◯	◯	◯	◯
Engelske bøker	◯	◯	◯	◯	◯
Andre engelske ting	◯	◯	◯	◯	◯

5. Har du vært i et engelsktalende land før? Og hvor lenge?

Markér bare én oval.

- ◯ Ja, mindre enn 3 måneder
- ◯ Ja, mellom 3-12 måneder
- ◯ Ja, mer enn 12 måneder
- ◯ Nei, aldri

6. Får du hjelp til å gjøre engelskleksene hjemme?

Markér bare én oval.

- ◯ Ja
- ◯ Av og til
- ◯ Nei
- ◯ Ikke hjemme, men jeg får hjelp av noen andre
- ◯ Jeg trenger ikke hjelp

7. Hvor mange år har du hatt engelsk undervisning?

Markér bare én oval.

- ◯ 7 år eller mer
- ◯ 4-6 år
- ◯ 2-3 år
- ◯ Under 1 år

8. Har du alltid gått på denne skolen?

Markér bare én oval.

- ◯ Ja
- ◯ Nei, jeg byttet skole etter 5 klasse
- ◯ Nei, jeg byttet skole før 5 klasse

9. Hvordan tror du at du har lært mest av engelsken du kan?

Markér bare én oval.

- ◯ Alt eller nesten alt gjennom skolen og lekser
- ◯ Det meste gjennom skolen og lekser
- ◯ Like mye på skolen og ved siden av skolen
- ◯ Det meste ved siden av skolen og lekser
- ◯ Alt eller nesten alt ved siden av skolen og lekser

Appendix 2: Table of group characteristics

Characteristics	Option	Control group	Experimental group
Gender	Boy	32%	50%
	Girl	54%	45%
	Do not want to answer	14%	5%
Speak Norwegian at home	Never/almost never	3%	2%
	Often/Almost always/always	97%	98%
Speak English at home	Never/almost never/rarely	75%	74%
	Often/Almost always/always	25%	26%
Speak another language at home	Never/almost never/rarely	82%	84%
	Often/Almost always/always	18%	16%
Speak English with friends or relative	Daily/weekly	30%	32%
	Monthly/yearly	20%	27%
	Never	43%	36%
Hours watching English movies/TV	More than 2 hours a day	24%	23%
	Less than 2 hours a day	63%	68%
	Never	13%	9%
Hours playing English games	More than 2 hours a day	40%	36%
	Less than 2 hours a day	53%	57%
	Never	7%	7%
Hours watching English YouTube	More than 2 hours a day	36%	20%
	Less than 2 hours a day	57%	75%
	Never	7%	5%
Hours spent on English social media	More than 2 hours a day	27%	18%
	Less than 2 hours a day	53%	68%
	Never	20%	14%
Hours spent reading English books	More than 2 hours a day	3%	0%
	Less than 2 hours a day	47%	36%
	Never	50%	64%

Hours spent doing other English activities	More than 2 hours a day	13%	11%
	Less than 2 hours a day	53%	57%
	Never	33%	32%
Been to an English-speaking country	Never	25%	19%
	Less than 3 months at the same time	68%	71%
	More than 3 months at the same time	7%	10%
Get help with homework	Yes/sometimes	61%	50%
	No	7%	10%
	I do not need help	32%	40%
Always gone to the same school	Yes	77%	77%
	No, changed after 5th grade	10%	2%
	No, changed before 5th grade	7%	16%
Where do you think you learned most of your English	Everything or almost everything from school of homework	17%	9%
	Mostly through school and homework	37%	46%
	Mostly from outside of school and homework	20%	34%
	Everything or almost everything from outside of school and homework	20%	5%

Appendix 3: Language attitudes and language anxiety survey + additional question

	Veldig uenig	Uenig	Delvis uenig	Litt enig	enig	Veldig enig
Jeg blir flau når jeg må svare på spørsmål i engelsktimen på engelsk						
Jeg liker å lære engelsk						
Jeg blir nervøs hvis jeg må snakke engelsk til en turist						
Det er ikke så viktig for meg å lære meg engelsk						
Jeg er rolig når jeg må snakke engelsk i timen						
Jeg hadde ikke likt å snakke engelsk på telefonen						
Engelsk er et veldig viktig fag på skolen						
Det bekymrer meg at andre i klassen er bedre til å snakke engelsk enn meg						
Det hadde gått helt fint hvis jeg skulle bestilt mat på engelsk						
Jeg er redd for at de andre i klassen skal le av meg når jeg snakker engelsk						
Jeg hater engelsk						
Jeg liker å snakke engelsk utenfor klasserommet						
Engelskundervisningen de siste to ukene har vært gøy (added only for posttest)						

Appendix 4: Debate booklet

Homework

Homework is something many pupils have to do after school. There has been a debate about homework for many years across many countries. You will now participate in this debate. You will be given a team and a side of the argument that you will present in two weeks.

Do not worry, we will prepare and be ready in time for the debate.

To have homework, or not to have homework, that is the question...

b

What is homework?

Homework, or a **homework assignment**, is a set of tasks assigned to students by their teachers to be completed outside the class. Common homework assignments may include required reading, a writing or typing project, mathematical exercises to be completed, information to be reviewed before a test, or other skills to be practiced.

Why are we debating?

Not everyone agree that homework is a good thing. Some schools have even stopped giving homework. Researchers are trying to figure out if homework is necessary or not, but they do not all agree.

The debate

It is now up to you to read and understand the debate. The material and articles that you can read in this paper will help you make arguments for the debate.

Remember to think about who wrote the sources and where they are coming from.

Remember to read both sides of the argument to be ready to answer the other side.

Articles and studies

A video from ABC:
https://www.abc.net.au/btn/classroom/homework-debate/11958930

Homework Debate

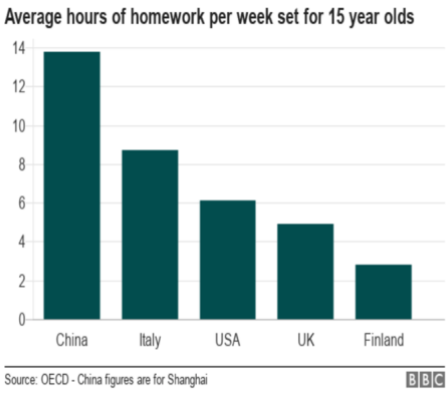
We take a look at the homework debate.

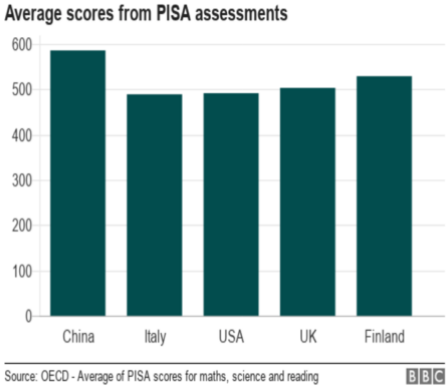
This is a video that explains the homework debate and some different views of the debate

Did you know that you can

CC

Statistics from an article from BBC:
https://www.bbc.com/news/education-43386670



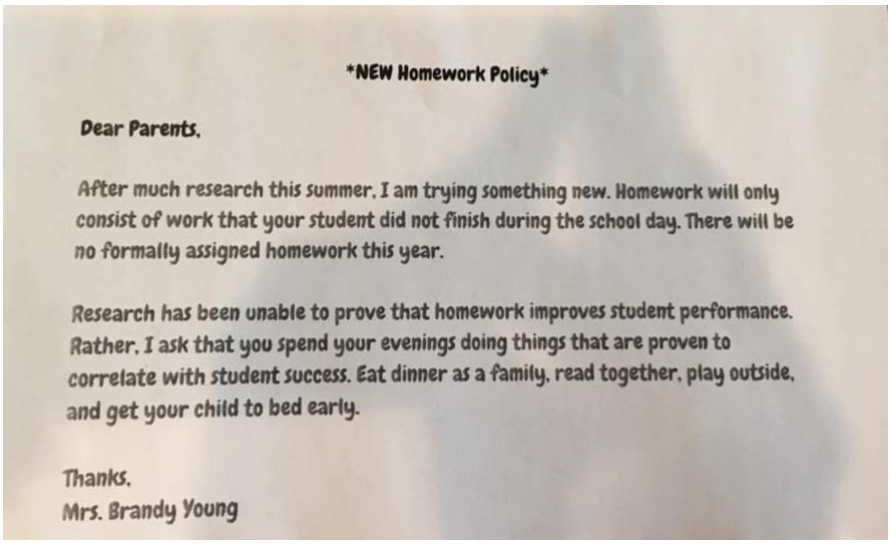


Pisa is a test that measures the skills of 15 years old's in mathematics, natural science and reading

TOP 14 reasons why homework is important:

1. It improves your child's thinking and memory
2. It helps your child develop positive study skills and habits that will serve him or her well throughout life
3. Homework encourages your child to use time wisely
4. It teaches your child to work independently
5. ….
   http://blog.eskool.ca/parenting/why-homework-is-important/

A Facebook post by an American teacher



NEW Homework Policy

Dear Parents,

After much research this summer, I am trying something new. Homework will only consist of work that your student did not finish during the school day. There will be no formally assigned homework this year.

Research has been unable to prove that homework improves student performance. Rather, I ask that you spend your evenings doing things that are proven to correlate with student success. Eat dinner as a family, read together, play outside, and get your child to bed early.

Thanks,
Mrs. Brandy Young

4

Is homework a good idea or not?

11 Jan 2017

Going to school - means lessons, assembly, seeing your friends and - for a lot of you - time to do homework!

While giving homework to pupils in secondary schools is generally seen as a good idea, some don't think that kids in primary schools should have to do it.

Why do people think homework is a good idea?

Many think that giving homework to primary school children is an important part of their learning. They believe it helps them to practice what that they have learnt in lessons, in order to get better at things like spelling and handwriting. They say it helps to teach children how to work on their own and be disciplined with themselves - both skills that are useful later in life. It can also allow parents or guardians to get involved in their children's learning.

To find out more about why people think homework is a good idea, Jenny spoke to Chris from the campaign for Real Education, which is a group of teachers and parents who care about how well schools are doing. Members of the organisation believe that traditional homework is important. Chris told Newsround: "If you like learning, homework helps to support your learning. It's really important to go back afterwards and think about what you're learning in class. Practice makes perfect."

"In parts of the world, children are doing much better in school than children in the UK. In most cases, they are doing much more homework."

"That doesn't mean you should be doing home work all the time."

"But a little bit of homework to support what you're doing in the classroom, involving your parents and guardians, is really good because it allows you to do as well as everybody else in the world."

Chris added that it is important to have a balance between homework and other activities. "Homework shouldn't be overdone. Let's do some homework and some play."

Why do people think homework is a bad idea?

Some people think that giving homework to children at primary school is not necessary. They think it puts too much pressure on them and that the time spent doing homework could be used to do other activities. Jenny also spoke to Nansi Ellis - assistant general secretary of one of the biggest teacher's unions in England, made up of teachers and heads - who doesn't believe that giving homework to primary school children is needed.

She told Newsround: "There is other good stuff you can do at home, like reading, playing sport or a musical instrument, or helping with the cooking, shopping or with your siblings. You might be a Guide or a Scout."

"Those things are really helpful for you to learn to work in a team, to learn to be creative, to ask questions and to help other people. These are really important skills."

"The trouble with homework is that it gets in the way of all of those good things that you could be doing and it doesn't necessarily help you with your school work."

GLOSSARY

- **Secondary school = ungdomsskolen**
- **Primary school = barneskolen**
- **Necessary = nødvendig**
- **Confusing = forvirrende**
- **Assigned = tildelt**

Sometimes parents or guardians try to help with homework and, if they have been taught differently, it can end up being confusing for the child doing the homework. They can also end up doing too much of the work themselves!

Nansi added: "Some children live in really busy houses with lots of people coming and going, and they don't have a quiet space to do homework, so they can't use it to help them to get better at studying on their own, which doesn't seem fair.

"Teachers set homework for you to get better at your learning - that seems like a really good reason. But actually, the evidence isn't clear that even that's true."

Another expert Rosamund McNeil, from a teachers' organisation called the NUT, said: "Pupils in Finland are assigned very little homework yet they remain one of the most educationally successful countries in the world."

https://www.bbc.co.uk/newsround/38383428

The different sides are introduced in videos if you follow this link.

Should homework be done in school time?

17 Oct 2012

Should homework be done in school time, rather than at home?

Might sound like a weird question but the President of France, Francois Hollande, thinks that it should. He's said he doesn't think it's fair that some kids get help with their work at home, when lots of others don't. So he wants to make the school day up to half an hour longer so that children can finish it there.

https://www.bbc.co.uk/newsround/19975538

Trust the kids: study shows they're great at homework if you leave them alone

Kathryn Lewis, May 22, 2018

Tell the grown-ups to back off — you can be trusted to do your homework without their help.

And you have an international study to back you up. It found children aged seven to 10 were better at homework if left alone.

The study's author, University of Eastern Finland Assoc Prof Jaana Viljaranta, said parents helping every step of the way risked making their children lose interest in schoolwork.

"One possible explanation* is that when the mother gives her child an opportunity* to do homework autonomously* the mother also sends out a message that she believes in the child's skills and capabilities*," Assoc Prof Viljaranta said.

"This in turn makes the child believe in him or herself, and in his or her skills and capabilities."

GLOSSARY

- **explanation: forklaring**
- **opportunity: mulighet**
- **autonomously: på egen hånd**
- **capabilities: evner/muligheter**

https://www.kidsnews.com.au/humanities/trust-the-kids-study-shows-theyre-great-at-homework-if-you-leave-them-alone/news-story/aa50b58a5c153b2e2c509bd0eca1663b

Do Kids Need Homework?

October 30, 2017

Yes!

By Janine Bempechat

Particularly for elementary school, homework has become a debated issue. That is because studies show that doing homework does not necessarily improve younger students' grades.

Still, many teachers and parents are in favor of homework in elementary school. They view it as a way for students to review what they are learning in school and develop the learning habits and study skills they will need through middle school and high school. Homework helps you practice how to plan your time, manage distractions, and persevere when learning becomes difficult.

Research shows that homework pays off. Students who invest effort in their homework are better at keeping track of their work and managing their time. When they begin middle school and high school, they are more positive about learning than students who do not invest time in homework. They also do much better in school.

This payoff may not appear during the elementary school years. But that doesn't mean homework should be abandoned. Instead, educators should focus on designing high-quality, enjoyable homework that is challenging in the best sense of the word.

No!

By Etta kralovec

I hated homework when I was in elementary school, and I hated it even more as a parent. As I recall, my kids and I argued more about homework than about anything else.

Now that I'm an educator, I spend a lot of time in schools. When it comes to homework, there is still a lot of controversy. And it's not just students who are unhappy. Many teachers and parents tell me they think homework is a waste of time. I couldn't agree more.

Homework has been around about as long as there have been schools. You may be surprised to know that it was not always popular. In 1905, California banned homework for students under 15. In the 1920s, doctors formed a group called Physicians Against Homework. They said homework harmed students' eyes and kept kids cooped up, denying them exercise and fresh air. More recently, researchers at Stanford University found that it causes considerable stress in students' lives.

My colleagues and I have put together healthy homework guidelines, which we hope will stimulate conversation between teachers, parents, and students, and eventually lead to some common-sense solutions for the age-old controversy of homework in our schools. It's a conversation worth having, don't you think?

https://www.timeforkids.com/g56/hwdebate/

GLOSSARY

- **Particularly - særlig**
- **Elementary school - barneskolen**
- **Persevere - bevare**
- **Payoff – lønne seg**
- **Abandon – forlate**
- **Recall – huske**
- **Harm – skade**
- **Cooped up – sperret inne**
- **Deny – nekte**
- **Considerable – en stor mengde**
- **Common-sense – sunn fornuft**
- **Controversy - uenigheter**

Useful phrases for debating

Building your Argument
Introducing your point: To begin with… First of all…
Connecting your points: Also… Furthermore… What's more…
Showing importance: More importantly… What's worse… Above all else…
Giving examples: For instance… For example…

Opinions, Preferences:
In my opinion..., The way I see it..., As far as I'm concerned..., If it were up to me..., I suppose..., I suspect that..., I'm pretty sure that..., I honestly feel that, Without a doubt…,

Disagreeing:
Don't you think it would be better..., Shouldn't we consider..., But what about..., I'm afraid I don't agree..., Frankly, I doubt if..., The truth of the matter is..., The problem with your point of view is that..., It's a fact that…, According to …, The reality of the situation is…, The numbers show that…, The fact is this:…,

Partially agreeing:
I agree with you to a point however…, I see where you are coming from but…,
I see what you are saying but …,

Delaying Strategies:
I can't answer that directly..., I'll need time to think about that..., That's a very interesting question, because…, That's a difficult question to answer,

Asking someone to repeat:
Pardon me? Pardon? Excuse me? Sorry? I'm sorry? I beg your pardon?

Expressing solutions and alternatives:
The solution is to…, Then you will…, The best way to … is …,
To …, you really have to …., There are many choices....,

Appendix 5: NSD application

Vil ditt barn delta i forskningsprosjektet *«Debate in the English Classroom»?*

Dette er et spørsmål til deg om ditt barn vil delta i et forskningsprosjekt hvor formålet er å se hvilken effekt det har å bruke debatt i engelskundervisningen. I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for ditt barn.

Formål
Formålet med dette forskningsprosjektet er å forstå hvilken effekt debatt har på elevers engelskferdigheter. Forskningsprosjektet vil foregå over to uker, hvor klassen til ditt barn vil delta. Det skal inngå i et masterprosjekt. For å ikke påvirke resultatene vil du ikke få vite nøyaktig hva prosjektet prøver å måle, dette kan du få informasjon om når prosjektet er ferdig på barnet ditt sin skole om du ønsker det.

Hvem er ansvarlig for forskningsprosjektet?
Høgskulen på Vestlandet, institutt for språk, litteratur, matematikk og tolkning er ansvarlig for prosjektet.

Hvorfor får du spørsmål om å delta?
Det er fire 7.klasser i Norge som skal delta på dette prosjektet. Klassene er valgt ut på grunn av personlig kjennskap til skolene hos student.

Hva innebærer det for deg å delta?
Hvis du velger at ditt barn kan delta i prosjektet, innebærer der at han/hun gjennomfører et spørreskjema og noen engelskprøver. Etter denne prøven og spørreskjema vil klassen bli delt inn i 2 grupper, disse gruppene vil vi beholde i alle engelsktimene i to uker. Den ene gruppen skal ha ordinær engelskundervisning og den andre gruppen skal ha engelskundervisning med student.

Hvis foresatte ønsker å ha innsyn i spørreskjemaet på forhånd er det bare å ta kontakt, så vil det sendes på mail.

Det er frivillig å delta
Det er frivillig å delta i prosjektet. Hvis dit barn velger å delta, kan du eller ditt barn når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle barnets personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg/dere hvis barnet ditt ikke vil delta eller senere velger å trekke deg. Hvis barnet ditt ikke ønsker å delta vil han/hun bli med gruppen som skal ha ordinær undervisning og han/hun vil ikke besvare hverken spørreskjema eller prøven.

Ditt barns personvern – hvordan vi oppbevarer og bruker dine opplysninger
Vi vil bare bruke opplysningene om barnet ditt til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.
Student, veileder og eventuelt en intern medarbeider vil ha tilgang til dataen vi samler.

Elevene vil bli tildelt et siffer som skal brukes på prøvene og spørreskjema, dette sifferet er det kun eleven som vet hva er. Datamaterialet vil kunne inneholde noen bakgrunnsopplysninger om eleven, men datamaterialet vil anonymiseres innen prosjektslutt og ditt barn vil ikke kunne gjenkjennes i endelig oppgave.

Prøvene og spørreskjemaene vil bli oppbevart på ark som oppbevares innelåst når det ikke er i bruk. I masteroppgaven vil barnet ditt sine besvarelser bli en del av en større statistikk og han/hun vil ikke kunne gjenkjennes i publikasjonen.

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?
Opplysningene slettes når prosjektet avsluttes/oppgaven er godkjent, noe som etter planen er 18.06.2021

Dine rettigheter
Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke personopplysninger som er registrert om deg, og å få utlevert en kopi av opplysningene,
- å få rettet personopplysninger om deg,
- å få slettet personopplysninger om deg, og
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger.

Hva gir oss rett til å behandle personopplysninger om deg?
Vi behandler opplysninger om deg basert på ditt samtykke.
På oppdrag fra Høgskulen på Vestlandet har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?
Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

- Høgskulen på Vestlandet ved Malin Hopøy (180241@stud.hvl.no) eller veileder, Matthew Scott Landers (Matthew.Scott.Landers@hvl.no).
- Vårt personvernombud: Trine Anikken Larsen, via email (personvernombud@hvl.no) eller via telefon (55 58 76 82).

Hvis du har spørsmål knyttet til NSD sin vurdering av prosjektet, kan du ta kontakt med:

- NSD – Norsk senter for forskningsdata AS på epost (personverntjenester@nsd.no) eller på telefon: 55 58 21 17.

Med vennlig hilsen

Matthew Scott Landers	Malin Hopøy
(Veileder)	(Student)

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet «*Debate in the English classroom*» og har fått anledning til å stille spørsmål. Jeg samtykker til:

- ☐ At mitt barn deltar i undervisningen
- ☐ At mitt barn svarer på spørreundersøkelsen og prøven
- ☐ At prøven og spørreundersøkelsen vil bli del av masterprosjektet «Debatt i klasserommet»

Jeg samtykker til at barnet mitt deltar i forskningsprosjektet «*Debate in the English Classroom»*

(Signert av foresatte til prosjektdeltaker, dato)

Appendix 6: NSD approval

NSD NORSK SENTER FOR FORSKNINGSDATA

NSD sin vurdering

Prosjekttittel

Debate in the English classroom

Referansenummer

793571

Registrert

19.10.2020 av Malin Hopøy - 180241@stud.hvl.no

Behandlingsansvarlig institusjon

Høgskulen på Vestlandet / Fakultet for lærerutdanning, kultur og idrett / Institutt for språk, litteratur, matematikk og tolkning

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Matthew Scott Landers, Matthew.Scott.Landers@hvl.no, tlf: 55587558

Type prosjekt

Studentprosjekt, masterstudium

Kontaktinformasjon, student

Malin Hopøy, malin.hopoy@hotmail.com, tlf: 40512619

Prosjektperiode

01.09.2020 - 15.05.2021

Status

12.11.2020 - Vurdert

Vurdering (1)

12.11.2020 - Vurdert

Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet

12.11.2020 med vedlegg, samt i meldingsdialogen mellom innmelder og NSD. Behandlingen kan starte.

MELD VESENTLIGE ENDRINGER
Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. Før du melder inn en endring, oppfordrer vi deg til å lese om hvilke type endringer det er nødvendig å melde:
https://nsd.no/personvernombud/meld_prosjekt/meld_endringer.html

Du må vente på svar fra NSD før endringen gjennomføres.

TYPE OPPLYSNINGER OG VARIGHET
Prosjektet vil behandle alminnelige kategorier av personopplysninger frem til 15.05.2021.

LOVLIG GRUNNLAG
Prosjektet vil innhente samtykke fra foresatte til behandlingen av personopplysninger om barna/elevene. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som foresatte kan trekke tilbake. Barna/elevene vil også samtykke til deltakelse.

Lovlig grunnlag for behandlingen vil dermed være foresattes samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a.

PERSONVERNPRINSIPPER
NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om:

- lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om og samtykker til behandlingen
- formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke viderebehandles til nye uforenlige formål
- dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet
- lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

DE REGISTRERTES RETTIGHETER
Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: åpenhet (art. 12), informasjon (art. 13), innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), underretning (art. 19), dataportabilitet (art. 20).

NSD vurderer at informasjonen som de registrerte og deres foresatte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13.

Vi minner om at hvis en registrert/foresatt tar kontakt om sine/barnets rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER
NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og eventuelt rådføre dere med behandlingsansvarlig institusjon.

OPPFØLGING AV PROSJEKTET
NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Lykke til med prosjektet!

Kontaktperson hos NSD: Marita Ådnanes Helleland
Tlf. Personverntjenester: 55 58 21 17 (tast 1)