Nøkkelord og Individuelle forskjeller – sett opp imot vokabularlæring og tekstforståelse

Keywords and Individual Differences – in relation to vocabulary acquisition and text understanding

Christopher Loe Olsen

Master in Education with English Didactics

Program for bachelor-masterstudier

Avdeling for Lærerutdanning

18.05.2016

ABSTRACT

Is using an explicit focus on vocabulary in the form of keyword identification an effective approach to help pupils understand a text about global warming and climate change, and can this focus on keywords facilitate pupils' acquisition of new vocabulary items? This Master's thesis explores several aspects related to explicit focus on vocabulary and individual differences between pupils, by using a quantitative approach. The individual differences explored are pupils' ability to infer word meaning from context, and their receptive vocabulary size.

A two-group quasi-experiment was conducted in a Norwegian lower secondary school. It examined the effect of using teacher-initiated focus on vocabulary on pupils' understanding of a text about global warming and climate change, and acquisition of new vocabulary items. The experiment group received teacher-initiated focus on vocabulary, and the control group received additional focus on the content, during two English sessions dealing with a text about global warming and climate change.

The data gathered during the quasi-experiment were analysed using independent means t-tests, Pearson's correlation, and simple linear regression modelling. The results indicate that the explicit focus on vocabulary was not more effective than focusing on the content of the text in relation to text understanding. However, the explicit focus on content appears to have caused a higher degree of acquisition of the vocabulary items in the text. A positive relationship between pupils' ability to infer word meaning from context and their vocabulary size was also found. This is in support of previous research. Finally, it was found that pupils' ability to infer word meaning from context accounted for a high amount of the variance in pupils' acquisition of new vocabulary items in the control group, but not in the experiment group.

In light of these results, I conclude that explicit focus on vocabulary in the second language classroom is important if the goal is to expand pupils' receptive vocabulary. Additionally, I make a strong case for the usefulness of this method in terms of helping pupils who are struggling with using contextual inferences of unknown words. Finally, I propose that some time should be spent on teaching pupils effective strategies related to contextual inferences of unknown words.

SAMMENDRAG

Er et eksplisitt fokus på vokabular i form av nøkkelord en effektiv måte å hjelpe elever med å forstå en tekst om global oppvarming, og kan dette fokuset på nøkkelord være med på å utvikle elevenes vokabular? Denne masteroppgaven utforsker flere aspekter ved det å bruke et eksplisitt fokus på vokabular og individuelle forskjeller via en kvantitativ metode. De individuelle forskjellene er elevenes evne til å forstå ukjente ord ut i fra kontekst, og deres vokabular størrelse.

Et kvasieksperiment med to grupper ble gjennomført i en norsk ungdomsskole. I eksperimentgruppen ble det brukt et eksplisitt fokus på vokabular, og i kontrollgruppen ble det brukt et økt fokus på innholdet i teksten. Det ble gjennomført to undervisningstimer i begge gruppene.

Dataene som ble samlet i løpet av kvasieksperimentet ble analysert ved hjelp av uavhengig t-tester, Pearsons korrelasjonskoeffisient, og enkel lineær regresjonsanalyse. Resultatene tyder på at det eksplisitte fokuset på vokabular ikke var mer effektivt enn fokuset på innhold i forhold til elevenes forståelse av teksten. Det eksplisitte fokuset på vokabular førte allikevel til at elevene lærte en større del av de ukjente ordene i eksperimentgruppen. Det ble også funnet et positivt forhold mellom elevenes evne til å forstå ukjente ord ut i fra kontekst og deres vokabular. Videre ble det oppdaget at elevenes evne til å forstå ukjente ord ut i fra kontekst forklarte en stor del av variasjonen i elevenes tilegnelse av nye ord i kontrollgruppen, men ikke i eksperimentgruppen.

I lys av resultatene konkluderer jeg med at et eksplisitt fokus på vokabular er viktig dersom målet er å utvikle elevenes vokabular. Videre argumenterer jeg for at denne metoden kan hjelpe elever som sliter med å forstå ukjente ord ut i fra kontekst. Avslutningsvis foreslår jeg at det bør brukes tid på å lære elevene effektive læringsstrategier rettet mot det å forstå ukjente ord ut i fra kontekst.

DECLARATION

This master's thesis was written as part of the master in education program at Bergen University College. All sources, including secondary sources, are cited in text, and added to the bibliography. The thesis consists of 34 518 words, including footnotes and references. This word-count does not include the appendix.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

4

5

CHAPTER 1 – INTRODUCTION

This master's thesis is inspired by the *Content and Language Integrated Learning* (*CLIL*) concept *language of learning*. It seeks to explore the possible effects of, and potential relationships between, using teacher-initiated explicit vocabulary teaching and two individual differences between pupils. This inquiry is conducted in an English as second language setting, in a Norwegian lower secondary school. The explicit focus on vocabulary is in the form of keyword identification (Appendix 1) during two English sessions. These sessions were oriented towards a text about global warming and climate change retrieved from the *National Geographic* education website (National Geographic, 2015). This is traditionally a topic encountered in the natural science and social studies subjects in the Norwegian educational system

The individual differences (*IDs*) explored are (1) *pupils' ability to infer word meaning from context* and (2) their *vocabulary size*. This investigation will try to expand our understanding of these IDs in second language (L2) acquisition by exploring the relationship between five variables. These variables are (1) *pupils' ability to infer word meaning from context*, (2) pupils' receptive vocabulary size, (3) *different teaching methods* (explicit focus on vocabulary vs focus on content), and the effect of these variables on (4) *pupils' understanding of the subject matter,* and (5) *acquisition of new vocabulary items*. This will be accomplished by using a quasi-experiment design with two groups.

The main justification for this exploration is practical. It is related to the context of the Norwegian educational system, particularly the principle of adapted education. Additionally, this inquiry will attempt to further improve our understanding of how individual differences play a role in the effectiveness of different teaching methods and activities.

1.2 Aims and scope of the thesis

In light of the recent trends of placing greater focus on professionalization of Norwegian teachers, research exploring questions that are related to classroom activities and teaching practices can be said to be highly needed. The relationship between research and teaching practices has undergone a number of changes, and the teacher is increasingly being viewed both as a researcher and consumer of research (e.g. Macaro, 2003, chapter 1; Taber, 2007, chapter 1). Thus, it is now possible to argue that contemporary educational research (and by extension L2 acquisition) should increasingly be focused on subjects that are relevant to

classroom practices. This does, of course, not mean that all research in the field of L2 acquisition should be classroom-based. However, it could be thought of as beneficial as this would make the research more approachable for practicing teachers.

In line with these new trends in educational research, this master's thesis seeks to accomplish two aims. The first aim is to explore the usefulness of the teacher-initiated identification and explicit focus on keywords as an approach to provide pupils' with an adequate understanding of the text about global warming and climate change (Natural Geographic, 2015). The term *teacher-initiated explicit focus on vocabulary* is defined as "any teaching activity initiated by the teacher and focusing on the meaning and usage of one or more words". This includes, but is not restricted to, the identification and presentation of keywords found in the text. Additionally, this also includes homework assignments featuring explicit focus on vocabulary. Since it is not inconceivable that pupils in the control group would use strategies involving explicit focus on vocabulary, such as the use of a dictionary, the prefix *teacher-initiated* is used.

This teacher-initiated explicit focus on vocabulary is inspired by the CLIL concept *language of learning*[1]. It is important to note that this focus on keywords in a text is not related to, or should be confused with, the keyword strategy first employed by Atkinson (1975), which is a mnemonic technique used to remember new words. Instead, the concept of keywords used during this investigation is defined as "potentially difficult words, that are likely important to understand in order to fully benefit a text and its content". In other words, keywords are associated with the lexical items in a text, and unlike *language for learning*[2] it does not extend to other areas of language such as syntax and morphology.

This focus on specific words in order to achieve adequate text comprehension is supported by for example Bauman, Edwards, Boland, Olejnik, and Kame'enui (2003, p. 450)

---

[1] CLIL is according to Coyle (2010, chapter one, 1st paragraph) "a dual-focused educational approach in which an **additional language** is used for the learning and teaching of both content *and* language,"
[2] Language of learning is defined as "an analysis of language needed for learners to access basic concepts and skills relating to the subject theme or topic" (Coyle, 2010, chapter 3, section 2.2, 5th paragraph).

who assert that "[i]f the goal is comprehension of a particular text, then teaching a specific set of words may be necessary to understand the text".

The second aim of this thesis is to explore the usefulness of this approach in relation to two individual differences. These differences are (1) pupils' ability to infer word meaning from context, and (2) their vocabulary size. Pupils' ability to infer word meaning from context should be understood as how adept an individual pupil is at the (meta)cognitive learning strategy related to the *use of contextual cues to infer the meaning of unknown words*. Vocabulary size refers to the individual pupil's *receptive vocabulary*. This aim is inspired by Skehan's (1989, chapter 1, section 1) assertion that there is a need for research into the relationship between individual differences, and the use of different teaching strategies and methodological approaches. Furthermore, in light of the principle of adapted education found in the quality framework for the Norwegian educational system, this aim is also related to the practices of Norwegian English teachers.

Both of these aims are related to two 10th grade LK06 competence aims, namely that pupils are expected to "[u]nderstand the main content and details of texts one has chosen", and to "[r]ead, understand and evaluate different types of texts of varying length about different topics"(UDIR, 2013). It is important to note that although the teaching method employed in the experiment group is inspired by the CLIL concept language of learning, this thesis does not explore the usefulness of the CLIL framework as such. There is no full immersion into the L2. Furthermore, although an authentic academic text using material normally associated with other subjects will be presented, none of the other defining characteristics of the CLIL framework will be present.

Additionally, as the text will cover topics of an academic nature, the aims of this thesis are related to the competence aim asserting that pupils are expected to "[c]ommunicate and converse about contemporary and academic topics" (UDIR, 2013). The communication in this case will be in the form of teacher–student interaction. Furthermore, communication will also be mediated through text reading. Thus, the aims of this thesis are closely linked to competence aims found in LK06. In order to accomplish these aims, four research questions will be explored:

Q-1. Does using teacher-initiated explicit vocabulary teaching in the form of identification and teaching of keywords found in the subject text provide a higher level of subject understanding in the experiment group compared to the control group?

Q-2. Does using explicit vocabulary teaching in the form of identification and teaching of keywords provide a higher degree of vocabulary item acquisition in the experiment group compared to the control group?

Q-3. Is there a relationship between pupils' ability to infer word meaning from context and their vocabulary size?

Q-4. Does pupils' ability to infer word meaning from context have an effect on the effectiveness of using teacher-initiated explicit focus on vocabulary in the form of keyword identification?

As these research questions are concerned with questions of cause and effect, the most appropriate research method is the quasi-experiment. The research carried out here is related to classroom activities and practices, and it is therefore not practically feasible to provide the random group assignment and confounding variable control needed to conduct a full-scale experiment. A two-group pre-test, post-test, delayed post-test quasi-experiment, with two dependent and three independent variables, was conducted in order to answer the four research questions. Thus, this thesis is predominantly quantitative.

The dependent variables were (1) acquisition of new vocabulary items (keywords), and (2) understanding of main concepts in the *National Geographic* (2015) text. The independent variables were one binary variable, and two numerical variables. The binary variable was group condition, with either explicit focus on vocabulary (experiment group, abbreviated *E*) or focus on content (control group, abbreviated *C*). The numerical variables were pupils' vocabulary size, and their ability to infer word meaning from context. The quasi-experiment was conducted over the course of five English sessions in both group conditions (10 sessions in total). Three of these sessions in both groups were used for the testing. Two English sessions were used to present and work with the text about global warming and climate change in both groups.

In order to answer the research questions, four hypotheses were formed.

H1: Using explicit vocabulary teaching in the form of identification and teaching of keywords will result in a higher understanding of the text about global warming and climate change in the experiment group compared to the control group.

H2: Using explicit vocabulary teaching in the form of keyword identification will cause a higher degree of keyword vocabulary items acquisition in the experiment group compared to the control group.

H3: There will be a positive relationship between pupils' ability to infer word meaning from context and their vocabulary size.

H4: The contextual inferences score on the pre-test will account for a high amount of the variance of the keywords score on the post-test in the control group, but not in the experiment group.

The first hypothesis (H1) postulates that the explicit teaching of vocabulary in the form of keyword identification will provide pupils with the necessary vocabulary items needed in order to understand the content of the text used during the teaching sessions. Thus, the pupils' understanding of the text measured on the post-tests should be higher in the experiment group, compared to the control group. The second hypothesis (H2) assumes that the explicit focus on vocabulary in the experiment group will cause the pupils to learn a higher number of the keywords compared to the control group. Because Paul Bloom (2000) proposed that literate people mostly learn new words by reading, the third hypothesis (H3) proposes that there will be a positive relationship between pupils' ability to infer word meaning from context and their vocabulary size. Finally, the fourth hypothesis (H4) asserts that pupils who scored high on the pre-test measuring their ability to infer word meaning from context will benefit less from the explicit teaching of vocabulary. I postulate that high-scoring pupils will be better equipped to learn and understand foreign words by trying to do inferences from context. Thus, pupils' ability to infer word meaning using contextual cues should be more important in the control group compared to the experiment group. If the score on the contextual inferences component of the pre-test functions as a predictor for the keywords score on the post-test in the control group but not in the experiment group, this hypothesis is strengthened.

1.3 Contribution to field

Hopefully, this master's thesis will be able to provide valuable insights that can be useful for Norwegian English teachers, by providing research that is directly linked to the competence aims found in LK06. If, for example, it turns out that using keyword identification in order to explore and understand academic texts can mitigate the effect of pupils' ability infer word meaning from context, this will have implications for how teachers use this method in order to comply with the principle of *adapted education* (see section 2.2).

It is a paradox that there is a principle in the Norwegian educational framework asserting that teachers must use different teaching methods and materials in order to accommodate the individual differences between pupils, when there according to Skehan (1989) is very little research into this area of L2 acquisition. It should be noted that Lightbown & Spada (2013, chapter 3) have identified several contemporary studies exploring the relationship between IDs and L2 acquisition. Hopefully, this thesis will also contribute to this rising trend in our understanding of how individual differences are related to different methodological approaches. Thus, this project should be in line with Skehan's (1989; 2002) call for such research. Finally, as the acquisition of new vocabulary items is explored in this investigation, a contribution may also be made to the field of L2 vocabulary research. This is important because Macaro (2003, p. 85) propose that "we need to investigate the teaching and learning conditions which promote a constant and rapid vocabulary growth".

1.4 Relationship to other studies

In 1989 Peter Skehan (1989, chapter seven) noted that there was a lack of studies exploring the relationship between learner IDs and language learning processes, in the field of L2 acquisition. He highlighted how researchers had made great leaps in our understanding of how learners acquire a second language, but asserted that a majority of the research up to that date had been oriented towards the similarities between learners of a second language instead of their individual differences. Skehan (1989, chapter 7, 1st paragraph) created the label *condition seeking research* to identify research into the relationship between individual differences and different teaching methods. In light of the research conducted up to that point, Skehan (1989) draws several important conclusions:

> "[…]first, that the completed condition seeking research has been some of the most fascinating in applied linguistics; second, that there are many studies whose

11

interpretation is not clear for research-design reasons, and third, that abysmal little research of this sort has actually been done". (Skehan, 1989, chapter 7, section 3, 1st paragraph)

As the relationship between two individual differences and different methodological approaches constitutes a major aspect of this investigation, it should be considered condition seeking research as defined by Skehan (1989). He identified four studies which explored these possible relationships. However, none of these studies explored the interaction between explicit focus on vocabulary, the ability to infer word meaning from context, and vocabulary size. In light of this, one can argue that this treatment–interaction relationship had been largely unexplored up to that point

A similar contemporary review of condition seeking research was conducted by Lightbown & Spada (2013, chapter 3). In relation to individual differences and classroom instruction, they assert that "there can be little doubt that an instructional approach that rigidly adheres to a single way of teaching all students and an expectation that all students can learn in the same way will deprive some students of learning opportunities" (Lightbown & Spada, 2013, chapter 3, section 3, 1st paragraph). They identify several studies suggesting that IDs are related to language learning success, and stress the importance of research into the relationship between IDs and L2 acquisition. Attitudes and motivation, intelligence, age, personality, and learning styles are identified as individual differences of interest in relation to L2 acquisition (Lightbown & Spada, 2013, chapter 3). Thus, individual differences and treatment–interaction relationships are now the subject of more research than during Skehan's (1989) review.

Several contemporary studies have explored different aspects of contextual inferences of unknown words, and these studies are presented in section 3.7.3 (e.g. Bolger et al., 2008; de Leeuw et al., 2014). Interestingly, a study by Bauman et al. (2003) explored several aspects of text understanding and vocabulary acquisition during teaching sessions related to social science. This included the effect of both instructions into how to use morphological and contextual cues to understand words, and vocabulary instruction. However, this study was conducted in a setting where English was the native language. Moreover, Nassaji (2006, p. 397) identifies that clear support for the importance of contextual inferences has not yet been found in an L2 setting. This study will try continue this work by exploring how different

methodological approaches are affected by, and related to, pupils' ability to infer word meaning from context, in an L2 setting.

1.5 Ethical considerations

Because this study investigates aspects of human behavior using young persons as participants, there are several ethical aspects that must be accounted for and taken into consideration before the research can start. Arguably, the most important ethical consideration is the retrieval of active informed consent. Because this study is being conducted on youths, this consent was retrieved from their parents using an informed consent form detailing what this study was going to investigate, and how the data was going to be used. This consent form can be found in Appendix 2. A second ethical consideration was ensuring anonymity for all of the participants in this study. Several steps were taken in order to achieve this, including, but not limited to, using pen and paper tests, and not recording any direct, sensitive, or triangulated (either though a key, or background information) personal information or characteristics. E-mail correspondence with NSD[3] was utilized in order to ensure this. One step in order to avoid any triangulation of personal information has been to not record the municipality or county, as well as the actual year level this study was conducted in. Because of this, the setting in which this study was conducted will be referred to as a Norwegian lower secondary school. In light of these steps, it is certain that the participants' anonymity has been ensured, and that no personal information has been recorded or handled during the data handling and analysis.

1.6 Structure

The structure of this thesis will be as follows: Chapter two will present important aspects of the Norwegian education system, and how this has affected the design of this study. Chapter three will present the theoretical background and previous research which has formed the basis for this inquiry. Chapter four will provide a detailed description of the quasi-experiment design, testing instruments, and the measures employed in order to ensure as high an amount of validity and reliability of the findings as possible. Chapter five will present the main findings,

---

[3] http://www.nsd.uib.no/personvern/

and how these are related to the four research questions that are explored. Finally, Chapter six will discuss the theoretical and practical implications of this study, and propose further areas of research, in light of the weaknesses and limitations of the quasi-experiment method.

CHAPTER 2 – BACKGROUND

This chapter outlines the background for the study. It presents relevant information concerning the Norwegian school system, with particular emphasis on the role of the core curriculum, quality framework, and subject curricula, and how these are related to English as a subject. Specifically, the viability of using topics and texts from other subjects, such as social science, during English sessions will be explored.

As this master's thesis explores the relationship between employing different teaching methods or approaches to teaching, it is according to Loewen & Philp (2012) related to the field of *instructed language acquisition* (ILA). They define ILA as "any systematic attempt to enable or facilitate language learning by manipulating the mechanisms of learning and/or the conditions under which they occur" (Loewen & Philp, 2012, section 1, 1st paragraph). Because this investigation requires the use of a substantial portion of a teacher's and two full classes' study time, it is important that the topics and methods used during this time are in accordance with the relevant curriculum. Thus, it is important to have a clear idea of what the pupils are expected to know and do during their education. This means, in a Norwegian context, having a good understanding of the knowledge promotion, as well as ensuring that the activities done in the classroom are in accordance with it (UDIR, 2013).

2.1 English in the Norwegian educational system

In the Norwegian educational system, the English subject is considered to be a crucial component. It is one of three subjects that require teachers to have 60 study points in lower secondary school, the other two being Norwegian, and mathematics. According to the knowledge promotion[4] (UDIR, 2013), English is part of the curriculum from grade 1 and onwards. It is part of a pupil's education until the end of grade 10, or even up to 3 in upper secondary school, depending on the study program. The subject of English involves not only the acquisition of a second language, in this case English, but it also encompasses the understanding of English-speaking countries' history, society and culture.

---

[4] Kunnskapsløftet LK06

15

Furthermore, the curriculum forming the foundation of the English subject in the Norwegian educational system highlights, on several occasions, the ability to use appropriate language and vocabulary, as well as the ability to understand the main content and details of different kinds of text. Thus, the ability to teach subjects of an increasing complexity using different sources of content matter can be considered highly important for an English teacher in the Norwegian educational system. It is therefore important to explore different approaches that can be employed in order to teach texts of varying complexity using an L2 (English), in the context of the Norwegian classroom.

One possible approach to achieve these competence aims is to use topics that have traditionally been taught in different subjects than English. This is not just a case of teaching these topics during English sessions instead of during other subject sessions, but also supplementing the pupils' understanding from the other subjects with English language. If a teacher carefully chooses the topics in such a way that they are recognizable from other subjects by the pupils, this might also enhance their ability to learn from the texts. Arguably, it is also important to employ different teaching strategies, such as teacher-initiated explicit focus on vocabulary, in order to ensure that all pupils are able to achieve an adequate understanding of the text employed to present these topics.

2.2 The Norwegian national curriculum – Core curriculum and quality framework

The Norwegian curriculum (LK06) for primary and secondary school can broadly be divided into three main sections. These sections are (1) subject curricula, (2) distribution of teaching hours per subject, and (3) the core curriculum and quality framework (UDIR, 2013). The first and last of these sections are highly relevant for this master's thesis, as the findings of the research conducted here can possibly have practical implications regarding these curricula. Thus, the first part of the background section will present the core curriculum and the quality framework, in order to discuss how these have affected the way this quasi-experiment has been designed.

> The core curriculum and quality framework for the Norwegian educational system is a continuation of the Education Act (UDIR, 2013). In the core curriculum the principles in the Education Act are expanded and presented in more detail. It presents the principal aims of education in Norway. (UDIR, 2005a, p. 5).

The core curriculum is divided into seven sections, stylized to highlight different "human beings". These human beings contain different areas of human development which the educational system is meant to encourage (UDIR, 2005a, p. 5). The seven human beings are called: The spiritual human being, the creative human being, the working human being, the liberally educated human being, the social human being, the environmentally aware human being, and the integrated human being. As these areas of human development are meant to be encouraged across all subjects, it is necessary to ensure that this is also done in the English subject.

One of the main justification for doing the classroom activities chosen in the quasi-experiment is to encourage the development of the pupils' environmentally aware human being. This is accomplished by choosing the topic global warming and climate change as the subject of the English sessions. In the core curriculum (UDIR, 2005a), different aspects related to our connection to and interaction with nature are presented. The effects and consequences of human industrial and scientific growth are highlighted and the section stresses the importance of ensuring a growth of the pupils' environmental awareness (UDIR, 2005a, p.35). According to the authors, this has implications for the classroom activities:

> The teaching must unite a clear understanding of nature's matter, forces and species with an appreciation of how social organization and technology both solve problems and impinge on the biosphere. It must spur the urge of the young to understand the processes of nature. (UDIR, 2005a, p. 38)

In light of this section of the Norwegian curriculum, it is clear that using the topics global warming and climate change in the English classroom is in accordance with the curriculum.

This thesis also examines the usefulness of teacher-initiated focus on explicit vocabulary in relation to individual differences between pupils. Thus, it might provide practical implications for how English teachers might have to plan teaching sessions in order to comply with the principles of adapted education and equal opportunities found in the quality framework. According to UDIR (2013):

> The quality framework helps to clarify the responsibilities the school owners (in Norwegian public schools the local and county administration authorities) have in providing education pursuant to the legislation and regulations and the principles of

17

human rights, and adapted to local and individual aptitudes, expectations and needs. (UDIR, 2005b, p. 1)

The concept of adapted education is highlighted and explained throughout the quality framework (UDIR, 2005b). The importance of using varied teaching methods, subjects, and structures in order to allow all pupils to thrive and learn in accordance with their personal aptitudes, interests and starting points, is the core principles of the concept of adapted education. The quality framework asserts that:

Adapted teaching for each and every pupil is characterized by variation in the use of subject materials, ways of working and teaching aids, as well as variation in the structure and intensity of the education. Pupils have different points of departure, use different learning strategies and differ in their progress in relation to the nationally stipulated competence aims. (UDIR, 2005 b, p. 5)

If the hypotheses in this inquiry are supported by the research data, this will have implications for how one as teacher should structure the teaching sessions in order to comply with the concept of adapted education. This will be explored further in section 6.3.

2.3 The Norwegian curriculum – subject curriculum English and social studies

Subject curricula of the Norwegian national curriculum are structured in such a way that they present a number of *competence aims* which the pupils are expected to have mastered after the year levels 4, 7, 10, and 1–3 in upper secondary school. In the introduction chapter of this thesis, several competence aims from the English subject curriculum relevant to the teaching used in the quasi-experiment were presented. These competence aims were:

- "Understand the main content and details of texts one has chosen."

- "Read, understand and evaluate different types of texts of varying length about different topics."

-"Communicate and converse about contemporary and academic topics."

All of these competence aims are from the year level 10 in the English subject curriculum. This means that the pupils are expected to have encountered teaching activities and topics related to these competence aims in year levels 8–10. They are all related to communication, and

more specifically the use of different texts. The third competence aim is, perhaps, the most important one in relation to this thesis. Communication in the case of the teaching activities used in this quasi-experiment will be both in written form (the subject specific text) and orally between teacher and students. The text about global warming and climate change (National Geographic, 2015) used for the teaching activities is arguably related to academic topics. Thus, the teaching activities chosen in this thesis are in accordance with the English subject curriculum.

However, the use of this topic is also supported in other areas of the Norwegian curriculum. As demonstrated in the previous section, the core curriculum highlights the environmentally aware human being as one of the areas of human development which Norwegian education is meant to encourage. Furthermore, the social studies curriculum contains several competence aims related to global warming and climate change. These competence aims are found in year level 7, and year level 10:

Year level 7

- "Explain relations between natural resources, industry, settlements and living conditions."

- "Explain how production and consumption can destroy ecosystems and pollute soil, water and air, and discuss and elaborate on how this might be prevented and repaired."

Year level 10

- "Investigate and discuss the use and misuse of resources, consequences this might have for the environment and society, and conflicts this can create locally and globally."

- "Explain the basic forces of nature focusing on internal and external forces on earth, movement in air masses, circulation of water, weather, climate and vegetation, and discuss and elaborate on the relationship between nature and society."

All of these competence aims are related to the topics explored during the teaching activities in the quasi-experiment. Furthermore, since some of these competence aims are from the year level 7, and the pupils in this thesis are from a lower secondary school, it is reasonable to

expect that they will have some background knowledge concerning global warming and climate change. Global warming and climate change are also topics that are commonly encountered in the media, and are arguably highly relevant to the pupils. This might have implication for the suitability of the text chosen for the teaching activities.

2.4 Summary

The sections 2.1 and 2.2 have presented the main areas of the Norwegian subject curriculum. They have demonstrated how the teaching activities used in the quasi-experiment are related to, and in accordance with, the core curriculum, as well as the English and social studies curricula. This is important because the quasi-experiment requires use of the pupils' and teacher's classroom time, and because this thesis is expected to provide insights that are of practical value to Norwegian teachers. If the topics and activities were not in accordance with the Norwegian national curriculum, the use of the teacher and pupils' time would be questionable. Furthermore, these sections have demonstrated how the principle of adapted education is related to H3 and H4 in this thesis. This will be explored further in section 6.3. Finally, section 2.3 demonstrated how the pupils can be expected to have encountered the topic used in the teaching activities during the course of primary school, as well as through media.

CHAPTER 3 – THEORY

This chapter will present the major underlying theories and research related to the relationship between vocabulary, individual differences between pupils, and reading comprehension. The reading comprehension aspect of L2 acquisition is closely related to and will be presented in conjunction with vocabulary size, with emphasis on the vocabulary needed for an adequate degree of comprehension. Other areas of reading comprehension will not be presented. The reference books *Teaching and Learning a Second Language* by Macaro (2003), *How Languages are Learned* (Lightbown & Spada, 2013), and *Researching Vocabulary – A Vocabulary Research Manual* (Schmitt, 2010) have functioned as the starting point for this theory chapter. These books have been further supplemented by contemporary research and more in-depth theoretical works related to the different areas of L2 acquisition.

The theory chapter is divided into two main sections. These sections are vocabulary and its relation to comprehension, and individual differences. The vocabulary section will present the major concepts and definitions used in contemporary vocabulary research, such as the terminology used to categorize vocabulary, differences between depth and size of vocabulary, and different levels word understanding. Additionally, this section will highlight the different ways new words are learnt or acquired, concentrating on the relationship between explicit and implicit focus on vocabulary. Finally, contemporary research related to vocabulary and its relevance for this thesis will be presented.

The individual differences section will highlight the individual differences between pupils. It will present how individual differences are related to second language acquisition, and how teachers plan their sessions. This section will contain an introduction to the three dimensions of language proficiency: fluency, accuracy, and complexity (FAC). Furthermore, this section will present research into the differences between pupils' L2 aptitude, with particular emphasis on grammatical sensitivity, and how this is related to L2 acquisition. Finally, the different strategies utilized by pupils to learn and understand new words, as well as their metacognitive abilities will be explored. Particularly, research related to how pupils can use contextual inferences of unknown words will be presented.

3.1 Vocabulary and its role in L2 learning

One can argue that words are one of the most important aspects of human language. Although words are made up of smaller units such as morphemes and phonemes, and language itself is

most certainly made up of larger units, words are a central unit for communication (Macaro, 2003; Schmitt, 2010, pp. 3–5). Macaro (2003, p. 85) cements the importance of words by asserting that "they are at the root of the processes we use for accessing language and they are probably the most important building blocks in the production of language".

The importance of an individuals' vocabulary can perhaps best be described by Macaro:

> So the first reason vocabulary is important in language learning is that words are complex things in relation to other words and to us humans that use them. Words imply and entail other words. Words may even trigger our though processes and therefore the utterances and sentences we produce. (2003, p. 62.)

Macaro further asserts that "as we know from educational studies in L1, the expansion of the lexicon is a key to educational success" (2003, p. 63). He explains that by expanding your vocabulary you can employ more complex strategies in order to understand new content. It will also make it easier to understand new words by guessing from context. This notion is also supported by Schmitt (2010, pp. 3–7), who further notes that a large vocabulary is a necessary component of communication. This will be further explored in relation to reading comprehension in section 3.3. Similarly, Macaro (2003, p. 63) asserts that a broader or larger vocabulary will facilitate better communication regardless of for example ones understanding of the grammatical rules of the L2.

Schmidt (2010, pp.3–7) highlights the often high degree of correlation between different aspects of vocabulary and other areas of language development, such as language proficiency. He presents several studies that have explored and demonstrated this relationship, and identifies the DIALANG tests as "one of the most systematical explorations of the relationship between vocabulary knowledge and language proficiency" (Schmidt, 2010, p. 4). According to Alderson (2005, p. 30) "[t]he DIALANG has tests in five aspects of language and language use: Reading, Listening, (indirect) Writing, Grammar and Vocabulary". Schmidt (2010, p.4) notes that the correlation between vocabulary and the other aspects of language was very high in the DIALANG tests. The highest correlation was .70 for the writing aspect. Schmidt (2010, p. 4) concludes that this data clearly supports his notion of the high importance of vocabulary.

According to Macaro (2003, chapter 4) vocabulary has been an increasingly researched area of L2 acquisition. He states that research suggests that both teachers and pupils consider vocabulary to be highly important in their learning of a foreign language. Even so, he notes that "vocabulary teaching as a separate and structured activity is remarkably lacking from foreign language courses whereas courses on the four language skills, on grammar and on the culture of the target language appear frequently in L2 courses" (Macaro, 2003, p. 63). He speculates that this is due to the increased focus on the impracticality of learning the high number of words and word families necessary in order to communicate in the L2, and the notion that learning vocabulary through exposure to new words is the most time effective way of acquiring a larger vocabulary. Finally, Macaro (2003, p. 86) stresses the importance of teaching the pupils effective learning strategies in order to expand their vocabulary and learn new words.

In the context of the Norwegian education system, this notion is supported by the lack of any explicit focus on English vocabulary in the competence aims after year level 4 (UDIR, 2013). In contrast, several competence aims for year levels (*YL*) 7 and 10 focus on teaching pupils effective learning strategies. These competence aims are found in YR 7 and 10. They state that pupils should "[i]dentify and use different situations and learning strategies to expand one's English-language skills" (YL 7), and "[u]se different situations, working methods and learning strategies to develop one's English-language skills" (YL10). This does, of course, not mean that there is no focus on vocabulary teaching activities in the Norwegian English classrooms. Nevertheless, it does provide evidence for Macaro's (2003) assertion.

One such potential learning strategy is the use of contextual cues in order to infer the meaning of unknown words (e.g. de Leeuw et al., 2014; Ellis, 1995). Macaro (2003, p. 86) asserts that "[v]ocabulary most certainly can be learnt implicitly, and context appears to be essential in ensuring in-depth understanding of a word". This will be further explored in section 3.6.3. In light of Macaro's statements, it is clear that there still is a need for research which explores the usefulness of teaching activities using explicit focus on vocabulary, adding to the volume of previous research identified by him. By focusing on the possible effect of teacher-initiated focus on vocabulary, both in terms of the understanding of the text and the learning/acquisition of new vocabulary items, this master's thesis is closely related to Macaro's statements.

23

3.2 Major terms used in vocabulary research

As this inquiry explores the relationship between explicit focus on vocabulary, vocabulary size, contextual inferences of unknown words, and the understanding of a subject text, it is necessary to have a clear understanding of the major terms and concepts used in vocabulary research. The terms *token*, *types*, *word families*, and *lemmas* will be presented and explained[5]. These terms are identified by Macaro (2003, pp. 63–64) as essential when dealing with vocabulary research. Furthermore, they are highly relevant to the selection criteria for the text about global warming and climate change used during the teaching sessions in this quasi-experiment. This is elaborated further in section 4.3.

The terms *tokens* and *types* are frequently used when exploring or analyzing texts. Tokens refer to the total number of words found in the text which is analyzed. Thus, one token refers to one word. For example, if a word count of a text shows 431 words, this means that there are 431 tokens in that text (Macaro, 2003, p. 63). This must not be confused with the term type. Type refers to the total number of different words in the text. Thus, a text can contain 431 tokens but unless all of these tokens are different words it will contain fewer types. In order to find the number of types in a text, it would therefore be necessary to find out how many words have been repeated and take away the repetitions from the total number (Macaro, 2003, p.63).

The terms *word families* and *lemmas* are superficially similar. According to Macaro (2003, p. 64) "a word family is made up of a "headword" (e.g. sustain) and all the words that are derived from it: sustainable, unsustainable, sustenance, sustainment". Lemmas are similar to word families but are according to Macaro (2003, p. 64) "more syntactically oriented". In the same way as a word family, a lemma also consists of one headword. However, instead of being made up of all the derivations of this word, a lemma contains all its inflected forms.

---

[5] Schmitt (2010, pp. 47-48) identifies an even higher number of terms in the field of vocabulary research, most of which are grounded in psycholinguistics. These terms are mostly related to the actual characteristics of a word (number of letters, syllables, morphemes, frequency of use, etc.) or its use (context, collocations, etc.) Although these terms are all important, they are considered to be outside of the scope of this master's thesis.

These terms are more closely connected to the difficulty of learning a new word, and all the words connected to it, as well as the ease with which one can guess the meaning of new words. For example, if a pupil knows the meaning and use of *sustain*, it will most likely be easier for him/her to guess the meaning of *sustainable*, or even *sustenance* (Macaro, 2003, p. 64).

It is common to organize words into different categories based on frequency (high/low) and area of use e.g. academic and technical (e.g. Coxhead et al. 2015; Macaro, 2003, p. 64). Function words are, for example, extensively used, and should therefore be understood as high frequency words (Macaro, 2003, p. 64). Academic words refer to words that are typically used in an academic context. It should be noted that academic words are also used in other types of text, but less frequently. Macaro (2003, p. 64) asserts that academic words are important to learn in an L2 context in order for pupils to be able to read and understand texts that deal with academic content (such as economics, natural science, or social studies). Considering the academic nature of the text dealing with global warming and climate change, it is likely that the pupils in this study will encounter some words that would appear on the lists of academic words.[6]

3.3 Vocabulary size and its relation to comprehension

As this master's thesis explores acquisition of new vocabulary items, it is necessary to have an overview of at which point a new word should be considered *known* or *learnt*. Furthermore, as the text used during the teaching sessions will be completely new for the pupils, and the themes explored in the text will most likely require the understanding of several academic

---

[6]It should be noted that Norbert & Diane Schmitt (2014) argue against the use of the division between high/low-frequency, academic, and technical words, in a pedagogical setting. They argue for a reassessment of the categories used to define the frequency of words. They propose that the threshold for high frequency words should be raised to the 3000 most commonly used word families and conversely that the threshold for low-frequency words should be lowered to the 9000 and less used word families. Furthermore, they identify the vocabulary in between these two categories as mid-frequency and argue for new research into this segment (N. Schmitt & D. Schmitt, 2014).

words, it is necessary to have an idea of the vocabulary size needed in order to understand authentic L2 texts. Additionally, one of the aspects explored in the pre-test of the quasi-experiment conducted in this investigation is the vocabulary size of the pupils. Finally, the post-test will measure the number of new words that the pupils have learnt during the course of the quasi-experiment. Thus, in order to use an appropriate instrument to measure the size of the pupils' vocabulary and their understanding of the keywords presented during the teaching sessions, it is necessary to have an overview of contemporary research related to these concepts.

Vocabulary size is a common measure used in both L1 and L2 vocabulary research. According to Coxhead et al. (2015, p. 122) "[a] test of vocabulary size measures how many words a learner knows" and "[i]t typically measures a learner's knowledge of the form of the word and the ability to link that form to a meaning". There are a number of different estimates for the size of English native speakers' vocabularies (cf. e.g Coxhead et al., 2015; Schmitt, Jiang & Grabe, 2011). It is important to keep in mind that Schmitt asserts that "[u]nfortunately much of the research into native speaker vocabulary size, has been methodically flawed, leading to wildly varying estimates" (Schmitt, 2010, p. 6).

In a cross sectional study using a Vocabulary Size Test (VST) on 227 New Zealand 13-18 year old native speakers of English, Coxhead et al. (2015) found a growth in vocabulary size by age group. They conclude that "[m]ost native speaking secondary school students are well advanced in their knowledge of the 9,000 high-frequency and mid-frequency words of English and teachers need to continue to support their vocabulary growth through their subject-matter study" (Coxhead et al. 2015, p. 133). Similarly, there are different estimates of the vocabulary size needed in order to properly use the L2 (e.g. Macaro, 2003 p. 65; Schmitt, 2010). Schmitt (2010) asserts that an L2 learner will not need to possess the same amount of words as a native speaker. This notion is also supported by Macaro (2003, p. 65) who stresses that different users of the L2 will differ in their use of the L2 language; thus, they will not need the same vocabulary size.

In relation to this master's thesis, the perhaps most important aspect of L2 vocabulary size is the number of words, or word families, needed in order to understand the text used during the teaching sessions. A common assertion is that it is necessary to understand 98% of the vocabulary in a text (cf. e.g. Nation, 2006, p. 78; Schmitt et al., 2011, p. 39). According to

26

Nation (2006, p. 79) "[i]f we take 98% as the ideal coverage, a 8,000–9,000 word-family vocabulary is needed for dealing with written text, and 6,000–7,000 families for dealing with spoken text". In light of this assertion N. Schmitt & D. Schmitt (2014) assert that:

> This vocabulary size takes us far beyond high-frequency vocabulary; in fact it takes us beyond current definitions of high-frequency, academic and technical vocabulary combined. If it takes this much vocabulary for proficient English use, there clearly needs to be a focus on vocabulary beyond that covered by the high-frequency, academic and technical categories. (N, Schmitt & D, Schmitt, 2014, p. 485)

In light of Macaro (2003) and Schmitt (2010), it is not possible to expect that the lower secondary school pupils in this study will have a native-like vocabulary size. Instead, various studies suggest that the number needed to function communicatively is substantially lower than that of a native speaker. For example, according to Coxhead et al. (2015, referring to; Schmitt et al., 2011):

> Typically, the research shows that the more words that a learner knows, the greater the likelihood of comprehension of the text (Schmitt et al. 2011), with a 98 % coverage of the vocabulary of the text being the minimum optimal level for most learners to gain adequate comprehension (Coxhead et al., 2015, p. 122).

Similarly, Macaro (2003, p. 65) identifies that it is necessary to understand 98% of the tokens in a text in order to read for pleasure. It should be noted here that the 98% figure appears to circulate a lot between different contemporary studies, and its exact origin is difficult to pinpoint. Furthermore, Macaro (2003, p. 65) also identifies a study by Laufer (1989) in which the number of tokens needed to be understood in order to understand the basics of a text is 95%. This assertion is also supported by Lightbown & Spada (2013, chapter 2, section 5). They additionally stress that that it will be difficult to infer the meaning of unknown words if less than 95% of the lexical items in the text are known. Thus, in light of these studies it appears that the number of known tokens in a text needed to understand the main content is between 95–98 percent. This assertion is also supported by Nation (2007, p.3).

3.4 Different levels of word knowledge

It is important to note that, although there has been a number of studies exploring the size of L1 or L2 speakers' vocabulary, measuring vocabulary size is not unproblematic. In their

27

introduction, Laufer & Goldstein (2004, pp. 401—402) assert that measures of vocabulary size alone can lead to a superficial measure of each language item, and that a measure of the depth of vocabulary can allow the researchers to explore several areas of word vocabulary knowledge. However, they also acknowledge that this form of vocabulary testing demands a lot of time. Additionally, they stress that "[s]ize tests allow for more efficient placement and admission in language teaching programs than depth tests or fluency tests" and that they "can function as relatively simple and efficient research instruments providing us with the vocabulary size of the participants at the beginning of the treatment" (Laufer & Goldstein, 2004, p. 402).

There are several views about what it means to know a word, and it is common to operate with different levels of word knowledge (cf. e.g. Coxhead et al. 2015; Laufer & Goldstein, 2004; Macaro, 2003; Schmitt, 2010; D, Schmitt & N, Schmitt, 2012). Such increments in word knowledge is one way to conceptualize vocabulary depth (Schmitt, 2010, p. 16). According to Macaro (2003, p. 68):

> Most researchers now recognize that these increments in word knowledge are related to the form of the word (what it looks and sounds like); the meaning of the word (all its semantic associations); the use of the word (how the word is used by the relevant language community).

Additionally, Laufer & Goldstein (2004) developed a bilingual vocabulary test, which in addition to measuring the vocabulary size, also measures different aspects of word knowledge. They propose four aspects of word knowledge, and assert that there is a difference in difficulty for achieving them. The four levels of word knowledge proposed by Laufer & Goldstein (2004, p. 399) are "passive recognition (easiest), active recognition, passive recall, and active recall (hardest)". In relation to this thesis, the tests used during the quasi-experiment would most likely fall within the levels of passive recognition and active recognition.

A common distinction in vocabulary knowledge research is that between receptive and productive vocabulary (Macaro, 2003, p. 65; Coxhead et al., 2015, p.122). The former can be understood as the words that a speaker can understand and recognize, whereas the latter can

be understood as the words which the speaker of a language is able to use and produce. In relation to the testing of different aspects of vocabulary, Coxhead et al. (2015) state that:

> A receptive test measures whether learners can provide or choose a meaning when they see the form of the word. A productive measure looks at the kind of knowledge needed for speaking and writing, so it measures whether learners can provide a word form to express a meaning. (Coxhead et al., 2015, p. 122)

In light of this definition, the vocabulary tests used in this inquiry will be related to receptive vocabulary knowledge.

3.5 How do we learn new words in an L2 setting? – Different vocabulary learning approaches and strategies

Having established that vocabulary is an important aspect of foreign language acquisition, and that there are different levels of vocabulary knowledge, it is also important to have an understanding of how new words are learnt. Is it possible to acquire a larger vocabulary through exposure only; are there any explicit methods that are more effective than others; is explicit focus on vocabulary necessary in an L2 learning context, and what learning strategies are available to the learners of new languages? (Macaro, 2003, p. 71). This section will provide an overview of these questions, and present relevant research into these aspects of L2 acquisition.

Macaro (2003, p. 71) proposes that vocabulary can be learnt either explicitly or implicitly, where explicit learning is the result of activities or strategies focusing directly on the words, whereas implicit learning occurs, in lack of a better word, automatically. This assertion is also supported by Ellis (1995, p. 3) who insists that most of the vocabulary that we know is learnt from context. A similar distinction is also proposed by Schmitt & Sonbul (2010) who assert that:

> Lexical researchers have proposed two main approaches to vocabulary learning in a L2: 'explicit learning', i.e. learning vocabulary when the focus is on the words to be learnt, and 'incidental learning', i.e. learning vocabulary as a by-product of any language learning activity, such as reading. (Sonbul & Schmitt, 2010, p. 253)

They further propose that in light of contemporary research, a combination of methods using explicit and incidental learning is effective (Sonbul & Schmitt, 2010, p.253). A similar

29

distinction is identified by Choo, Tan, and Pandian. (2012, p. 853), who stresses that the deciding factor is the "presence or absence of conscious processes". They also insist that both explicit and implicit vocabulary learning should be part of the L2 classroom (Choo et al. 2012, p. 857). A different distinction is proposed by Ellis (1995) who distinguishes between implicit and explicit vocabulary acquisition. This is explored further in section 3.6.3.

There has been a number of studies showing that it is possible to develop different aspects of the L2 using implicit teaching approaches leading to incidental learning. One such approach is extensive reading. In 1983 Elley & Manghubai (1983) conducted a vast experimental study in rural Fiji, exploring the usefulness of using two different variants of a new technique called a *book flood* on pupils' English language development compared to a control group. They believed that the L2 could be developed using methods found in L1 teaching, and wanted to provide empirical evidence supporting this (Elley & Manghubai, 1983, p. 54). They found that the experimental groups outperformed the control groups in several areas of language development (Elley & Manghubai, 1983, p. 65). Although there were some shortcomings in this study (mainly related to the tests used and degree of experiment condition control, it is today considered one the breakthrough studies in the field of L2 acquisition (Drew & Sørheim, 2009, p. 77).

Since Elley & Manghubai (1983) presented their findings, several studies have explored the usefulness of extensive reading (e.g. Chun et al., 2012), and the usefulness of explicit teaching strategies versus implicit learning (e.g. Schmitt & Sonbul, 2010). Chun et al. (2012) conducted a two group 9 week experiment (pre-test, post-test, delayed post-test design) contrasting the effectiveness of extensive reading (*ER*) and paired-associate learning (*PAL*) on vocabulary learning. Using a combination of ANOVA and t-test analyses, they found that both groups had learnt a significant amount of new words. Furthermore, the two groups did not differ significantly on the pre-test and post-test scores. However, on the delayed post-test scores the ER group showed a greater amount of retention compared to the PAL group. In light of their findings, Chun et al. (2012, p.128) conclude that both methods were effective in promoting the learning of new words, and that extensive reading was "considerably more effective than PAL in long-term vocabulary retention".

Sonbul & Schmitt (2010) conducted an experimental study exploring the efficiency of using explicit vocabulary techniques after reading compared to incidental learning through

30

reading only, using 40 female medical students. A 700 word text was chosen, and low-frequency medical words were selected from this text and assigned to two different conditions: read only, and read-plus (word is read, plus explicit instruction). This was a one group (all the students were exposed to the reading only and reading-plus explicit vocabulary words) pre-test, post-test and delayed post-test design. The post-test measured three areas of word knowledge. These were form recall, meaning recall, and meaning recognitions. Sonbul & Schmitt (2010, pp. 256–257) found that the read-plus condition resulted higher scores in the delayed post-test compared to the read only condition. In light of these findings they conclude that incidental learning of new words did occur, but that the explicit focus lead to greater and deeper learning (Sonbul & Schmitt, 2010, p. 258). Furthermore, they identify some practical implications in light of these results:

> Leaving L2 students to learn vocabulary in context does not seem to get them any deeper than meaning recognition knowledge (i.e. the ability to recognize word meanings on a multiple-choice test). In any vocabulary teaching programme [sic], where the purpose is achieving deeper levels of vocabulary knowledge, direct instruction should be adopted. (Sonbul & Schmitt, 2010, p. 258)

One study attempting to examine the effect of supplementing incidental word learning (in the form of contextual factors), with explicit focus on vocabulary (in the form of definitions) was conducted by Bolger, Ballas, Landen & Perfetti (2008). They postulated that although contextual factors are important in learning new words, contextual factors can be enhanced by providing the pupils with definitions. They examined two hypotheses over the course of two experiments. These hypotheses were that "variation in contexts is important for allowing core meaning features of a word to emerge" and that "definitions are effective because they can interact with contexts to communicate core meanings" (Bolger, et al, 2008, p. 124).

In light of their findings, they draw several important conclusion that are relevant to this master's thesis. Perhaps the most important conclusion drawn by Bolger et al. (2008, p.145) was that although the effect of using definitions was limited, they are still valuable additions to simply seeing words in context. Instead of viewing these two approaches to vocabulary as opposed to each other, Bolger et al (2008, p.145) propose that they should be used together in order to provide the highest amount of vocabulary learning. They propose that "[t]he value of sentence contexts and definitions are complementary in the following

31

sense: Sentences provide referentially specified predication for a new word, whereas definitions add pointers to meaning boundaries" (Bolger et al, 2008, p 145). This proposal has had a large influence on how the keywords have been presented and worked with in the experiment group in this inquiry.

Bauman et al. (2003) conducted an extensive study in an L1 context. They examined the effect of either teaching pupils how to use morphological and contextual cues (*MC*) in order to infer word meaning, or textbook vocabulary instruction, on several aspects of 5th grade pupils' learning. This included acquisition of new vocabulary items, and understanding of social studies content. They found no difference between the groups in relation to the social studies content understanding (Bauman et al., 2003, p. 486). However, they found that the textbook vocabulary group outperformed the MC group on learning new vocabulary items (Bauman et al., 2003, p. 480). In light of their findings, Bauman et al. (2003, pp. 487–488) conclude that instructions into using morpheme cues to infer word meaning was effective. However, this does not extend to the use of contextual cues, which only had a limited effect.

Finally, a study by Zaid (2009) using a two-group quasi-experiment with 34 participants (17 in each group) explored the research question "which is more effective for the presentation of vocabulary for Arabic-speaking EFL learners: vocabulary presentation in context or in isolation? (Zaid, 2009, p. 58). He found that both group conditions had resulted in an increase in the group's vocabulary.

These four studies represent a minor portion of all the studies that have been conducted in relation to different approaches to vocabulary teaching/learning. Even so, it appears that both explicit focus on vocabulary and incidental learning of vocabulary can be effective in an L2 setting. In relation to this master's thesis, it is clear that any learning of new vocabulary in the control group will be a result of implicit or incidental learning as described by Schmitt & Sonbul (2010) and Bolger et al. (2008). In the experiment group, any vocabulary item learning will instead be a combination of explicit focus on vocabulary and seeing this vocabulary in context.

3.6 Acquisition of vocabulary – How does it occur?

In light of the research that has been conducted into the effectiveness of incidental learning of vocabulary, it is clear that this approach to working with vocabulary has shown promising

results (cf. e.g. Chun et al. 2012; Manghubai & Elley, 1983; Schmitt & Sonbul, 2010). This suggests that incidental learning of new vocabulary has a place in the L2 classroom. It should be noted that Macaro (2003, p. 86) identifies that although vocabulary can be learnt incidentally, "[t]he research evidence is quite strong that inferred vocabulary is enhanced by explicit and methodological vocabulary learning". Having established this, it is necessary to have an understanding of the underlying mechanisms of vocabulary acquisition. Two competing hypotheses will be explored in this section as possible explanations of vocabulary acquisition[7]. These hypotheses are Krashen's (1982) input hypothesis, and explicit language acquisition (Ellis, 1995).

3.6.1 Krashen's Principles of Second Language Acquisition

One of the major proponents of incidental learning of vocabulary is Stephen D. Krashen (1982). His work, *Principles and Practice in Second Language Acquisition,* is considered one of the seminal works in the field of L2 acquisition. It is important to note that this book was launched one year before Manghubai & Elley's (1983) study was published, and in light of the audio-lingual methodology which was still prevalent at this time, Krashen's (1982) ideas were (and still are) controversial. Krashen (1982) proposed that the principles and processes that underlie the development of a first language during childhood are, to some extent, applicable to the learning (or more correctly the acquisition) of a second language. He defines acquisition by asserting that "[l]anguage acquisition is a subconscious process; language acquirers are not usually aware of the fact that they are acquiring language, but are only aware of the fact that they are using the language for communication" (Krashen, 1982, p. 10). By applying this principle to vocabulary, vocabulary acquisition in light of Krashen's (1982) theory is a subconscious process where the learner is not aware that he/she is expanding his/hers vocabulary.

It should be noted that the principles of Krashen's theory are related to areas of language acquisition and learning that are well beyond the scope of this thesis, such as syntax

---

[7] For the purpose of this thesis the distinction between vocabulary learning and vocabulary acquisition is that vocabulary acquisition happens as a byproduct of other activities such as extensive reading. Conversely, vocabulary learning occurs as a result of explicit vocabulary learning activities.

and grammar. Thus, it will not be possible to present this theory in too much detail. Instead, the focus will be on how vocabulary acquisition occurs in light of Krashen's principles.

3.6.2 The input hypothesis

A major aspect of Krashen's (1982) theory of how language, and by extension vocabulary, is acquired, is the input hypothesis. He formally defines this hypothesis by stating that:

> The input hypothesis makes the following claim: a necessary (but not sufficient) condition to move from stage i to stage i + 1 is that the acquirer understand input that contains i + 1, where "understand" means that the acquirer is focussed [sic] on the meaning and not the form of the message (Krashen, 1982, p. 21).

More informally he postulates that we only "acquire" when we are presented with, and exposed to something that is a little beyond our current level. In order for the input to result in any acquisition, Krashen (1982, p. 63-71) asserts that it must be (1) comprehensible, (2) interesting and/or relevant, (3) not grammatically sequenced[8], and (4) in sufficient quantity. In light of Krashen's theory (1982), it is clear that the input used during the teaching sessions in the investigation carried out here should conform to these characteristics. The notion of comprehensible input can also be thought of as being in line with the principle of using texts with 95-98 % known vocabulary as identified in section 3.3.

3.6.3 Competing hypotheses: explicit acquisition

Although Krashen's work has had a major impact on how we now believe that second languages can be acquired rather than learnt, his principles are not without criticism. Ellis (1995) proposes that one of the underlying principles of the input hypothesis is to some extent incorrect; one of the implications of the input hypothesis is that the acquisition of language, including vocabulary, is subconscious. In other words, the person who is acquiring the language is not aware of this process. In the case of unknown vocabulary, the acquisition

---

[8] This means according to Krashen (1982, p. 68) that in "acquisition-oriented materials, we should not be consciously concerned about including i + 1 in the input. Part (3) of the Input Hypothesis claims that when input is comprehensible, when meaning is successfully negotiated, i + 1 will be present automatically, in most cases".

according to Krashen's input hypothesis would happen through exposure alone, without the reader noticing the unknown word (Krashen 1982, Ellis, 1995).

Ellis (1995, pp. 5–6) proposes that this hypothesis is flawed, or more correctly too narrow. He does not contradict Krashen's (1982) input hypothesis, but instead proposes a second hypothesis suggesting that humans are capable of acquiring new vocabulary both implicitly and explicitly. He maintains that acquisition is not exclusively passive, and that the acquisition of new vocabulary can become more effective by using a number of metacognitive strategies. Examples of such metacognitive strategies, according to Ellis (1995, p. 5), are:

> […](i)noticing that the word is unfamiliar, (ii) making attempts to infer the word from context (or acquiring the definition from consulting others or dictionaries or vocabularies), (iii) making attempts to consolidate this new understanding by repetition and associational learning strategies such as semantic or imagery mediation techniques.

This must not be misunderstood as teacher lead explicit vocabulary *learning* activities. Instead, Ellis postulates that although we acquire the majority of our vocabulary through for example reading, it does not mean that the reader is not actively working with the unknown vocabulary. This notion is also supported by Ghanbari & Marzban (2014, p. 3855) who propose that learning strategies can play an important role in L2 development, because they encourage the learner's active involvement in the learning process.

3.7 Individual differences

This section will focus on the unique set of abilities, aptitudes, and interests that constitute individual differences between pupils. In order to set this field apart from other fields focusing on other forms of personality traits, the study of these forms of individual differences has been labeled differential psychology (Larsen & Buss, 2008, p. 97). In essence, every single aspect of human personality, experience, knowledge, etc., can have an effect on how successfully a new language or area of language is acquired, and which methods and teaching activities is best suited for the individual pupil (Skehan, 1989).

Skehan (1989) and Robinson (2002) suggest that individual differences affect each other. Because of the vast expanse of different individual differences which can be proposed as important for L2 acquisition, it will not be possible to give an in-depth presentation of all

35

possible individual differences. The immediately ensuing section will present several individual differences which can have an effect on the use of different teaching methods and activities. These individual differences are related to motivation, language proficiency (fluency, accuracy, and complexity), language aptitude, and contextual inferences of unknown vocabulary. Because contextual inferences of unknown vocabulary is the main individual difference explored during this quasi-experiment, the man part of this section will be devoted to this individual difference. This includes research into how it has been shown to have an effect on different teaching activities between pupils (de Leeuw, et al, 2014).

I have created Figure 1 to illustrate the multitude of components which can be thought of as individual differences affecting L2 language development. I have classified these individual differences as either static, dynamic, or expanding. It should be noted that this is not a conclusive list of IDs that can affect L2 development.

Static	Dynamic	Expanding

• Language Aptitude
 	
• Intelligence
  • IQ
  • Working Memory
 	
• Verbal Ability
	• Personality
	  • Motivation
	  • Locus of controll
	  • Efficacy Expectations
	  • Outcome Excpetations
	  • Self-Efficacy
	  • Arousal
	
	• (Meta)cognitive Strategies
	  • Contextual Inferences
	    • Local Inferences
	    • Global Inferences
	  • Mnemonics
	    • Keyword Strategy
		• Proficiency
		  • Fluency
		  • Accuracy
		  • Complexity
		
		• Procedural Knowledge
		  • Syntax
		  • Morphology
		
		• Vocabulary
		  • Size
		  • Depth

Figure 1 - *List of potential individual differences affecting L2 acquisition.*

Individual differences identified as static are thought of as aspects of the induvial pupil which are relatively constant. For example, language aptitude is often thought of as a stable individual difference. It is often measured by some form of test battery (e.g. MLAT, Canal-F).

36

To illustrate, the test scores on the MLAT testing battery should according to Lltf.net (2011, p. 4) be "considered to be valid for at least 5 years" and that "language aptitude as measured by the MLAT is viewed as a stable trait, one not readily subject to change or improvement through courses or experience". In light of this, although the components of the MLAT testing battery might not in themselves be stable over time, the score indicating the level of *language aptitude* could be considered a stable individual difference.

Dynamic differences are differences which can change freely over time, regardless of direction. For example, the use of different metacognitive strategies can change as the pupil either adopts new metacognitive strategies, or chooses to use the most appropriate one in any given L2 situation. Similarly, motivation can change over time or as a result of different conditions. For example, according to Reeve (2009, p. 374) arousal can affect alertness, wakefulness and activation, and affects motivation in several ways. It is affected by environment stimulation, and people seek behaviors that either increase or decrease their arousal. To illustrate, Reeve (2009, p. 374) suggests that people perform best when they are optimally aroused. Additionally, locus of control, efficacy and outcome expectations, and self-efficacy can affect how individuals act (e.g. Larsen & Buss, 2008; Reeve, 2009; Skaalvik & Skaalvik, 2005).

According to Reeve (2009, p.231) "an efficacy expectation is a judgement of one's capacity to execute a particular act or course of action", whereas "an outcome expectation is a judgement that a given action, once performed, will cause a particular outcome". Locus of control is defined by Larsen & Buss (2008, p. 688) as "a concept that describes a person's perception of responsibility for the events in his or hers life". They stress that efficacy expectations affect the chance of for example a pupil doing a certain action. Finally, self-efficacy is defined as "one's judgement of how well (or poorly) one will cope with a situation, given the skills one possesses and he circumstances one faces" (Reeve, 2009, p. 233, referring to; Bandura, 1986, 1993, 1997). Self-efficacy is affected by several sources. These sources are "personal behavior history, vicarious experience, verbal persuasion (pep talk), and physiological activity" (Reeve, 2009, p. 238). The importance of self-efficacy, and efficacy and outcome expectations, in regulating behavior is stressed by Skaalvik & Skaalvik (2005, pp. 146–148). They assert that pupils who have positive efficacy expectations tend to use more appropriate learning strategies compared to other pupils.

37

Finally, areas of L2 development such as proficiency or vocabulary have been classified as expanding, signifying that these differences are (hopefully) being improved and expanded over time.

3.7.1 Language Proficiency

Arguably, a logical starting point for any study exploring the concept of individual differences is the individual pupil's level of L2 proficiency. As was noted in section 2.2, the concept of adapted education demands that each pupil's individual starting point must be taken into consideration when the teaching is being planned. As the "starting point" of every pupil is a loose term, it could be thought of as their level of L2 proficiency. However, what constitutes proficiency has not been agreed upon until recently (Housen, Kuiken, Vedder, 2012, p. 1). Today there is a general agreement that language proficiency has three main dimension: fluency, accuracy, and complexity (FAC). Some of these dimension, notably fluency, can further be divided into subdimensions (Housen et at, 2012, pp. 1–2). Additionally, how these dimensions of L2 proficiency can and should be operationalized in such a way that they can consistently be subjected to research is still not clear (Housen et al. 2012, p. 8).

Fluency can according to Housen et al. (2012, p. 5) be understood either in a narrow or more general sense. They assert that that although fluency has traditionally been thought of as how readily a user of the L2 can access and produce the L2 in written or oral form, it is by contemporary researchers now defined much more narrowly. Today, it is thought of as consisting of at least 3 subdimensions. Housen et al. (2012, p. 5) identifies these subdimensions as "speed fluency (rate and density of linguistic units produced), breakdown fluency (number, length and location of pauses), and repair fluency (false starts, misformulations, self-corrections and repetitions)". Finally, they conclude that in light of these new narrow definitions of fluency, fluency is today more oriented towards oral language production (Housen et al., 2012, p. 6).

Accuracy is according to Housen et al. (2012, p. 4) understood as the degree to which the language produced conforms to the rules and norms of the L2. They identify that although this has been thought of as one of the more agreed upon dimension of proficiency, there is some disagreement about what should be considered as "errors" or deviations from the norm. They argue that the term accuracy should be replaced, or supplemented with, either "appropriateness or acceptability" (Housen et al, 2012, p. 4).

38

Finally, complexity is identified as the most debated aspect of L2 proficiency. It is by Housen et al. (2012, p. 4) identified as having two broadly different definitions. They assert that the "term is used in SLA literature in at least two different ways: as linguistic complexity and as cognitive complexity" (Housen et al, 2012, p. 4). Cognitive complexity is defined by Housen et al. (2012, p. 4) as "the relative difficulty with which language elements are processed during L2 performance and L2 learning". Conversely, linguistic complexity is defined by Housen et al. (2012, p. 4) as "an objective given, independent from the learner, which refers to the intrinsic formal or semantic functional properties of L2 elements (e.g. forms, meanings, and form-meaning mappings) or to properties of (sub-) systems of L2 elements".

Housen et al. (2012, p. 6) identify two major competing models of the interactions between elements that make up language proficiency. These models are the limited attentional capacity model, and multiple resources model. The former postulates according to Housen et al. (2012, p. 6, referring to; Skehan, 2009) that "humans have a limited information processing capacity and L2 learners must therefore prioritize where they allocate their attention during task performance, so that attention allocated to one dimension of language production will be lost on others". The latter instead proposes that attention pools for the different components of proficiency can be drawn upon independently (Housen et al., 2012, p.6, referring to; Robinson, 2001b; 2005).

Figure 2 presents a model of language proficiency development proposed by Housen et al. (2012, p. 6):
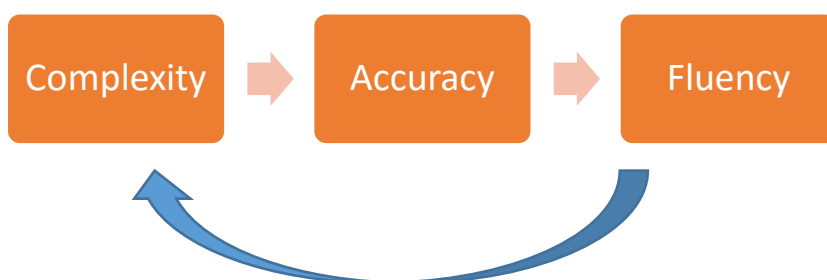
Figure 2 - *Proposed model of L2 proficiency development (Housen et al., 2012, p. 6).*

This model suggests that L2 proficiency is developed by the learner acquiring new and more complex language structures. When these new structures are being internalized, greater

accuracy is developed. Finally, as the new structures are familiarized greater fluency is achieved (Housen et al., 2012, p. 6).

Even though language proficiency is an important individual difference, the suitability of using different teaching methods on pupils with different levels of language proficiency has not been documented, to my knowledge. One of the reasons for this is most likely the issues with testing and operationalization identified by Housen et al. (2012). In relation to this master's thesis, it would most likely be too ambitious to make an attempt at measuring the language proficiency of two secondary school classes. Nevertheless, the potentially positive effect of using teaching methods adapted to the level of L2 proficiency for each pupil should not be ignored. This is explored further in section 6.4.

3.7.2 Language Aptitude

Language aptitude is an individual difference which has received an enormous amount of attention (Skehan, 1989, chapter 3; 2002). Different testing batteries attempting to predict the success of language learning have been developed, and are in use today. Perhaps the most well-established of these batteries is the Modern Languages Aptitude Test (MLAT) developed by J. B. Carroll[9]. According to Skehan (1989) this test measures four areas of language aptitude which are thought of as being stable, and not prone to trainability:

1. Phonemic Coding Ability
2. Grammatical Sensitivity
3. Inductive Language Ability
4. Rote learning activity for foreign language materials (Skehan, 1989, chapter 3, section 2, 4th paragraph)

Of the areas of language aptitude, grammatical sensitivity has been shown to be the best predictor of language learning success. It is defined by Skehan (1989, chapter 3, section 1, 6th paragraph) as the "ability to recognize the grammatical functions that words fulfil in sentences".

---

[9] Other language aptitude testing batteries include the Canal-F (Grigorenko, Sternman & Ehrman, 2000).

40

However, the importance of grammatical sensitivity has recently been contested by Vanpatten et al. (2013) and Vanpatten & Smith (2015). Based on their four experiments exploring the effect of explicit information and grammatical sensitivity on the learning of the first noun principle, Vanpatten et al (2013, p. 506) suggest that their "results also show that grammatical sensitivity does not significantly correlate with any of the measures under any conditions except for the German group that received explicit information". In light of this they conclude that "in terms of grammatical sensitivity, the results of the present experiments do not support the more general claim that grammatical sensitivity is a significant variable in SLA" (Vanpatten et al., 2013, p. 524).

Similarly, in light of their research Vanpatten and Schmitt (2015) challenge the importance of the grammatical sensitivity component of the MLAT testing battery. They contest two major expectations in language acquisition research:

"Namely, we challenge (a) the popular assumption that language consists primarily of rules and (b) its corollary, that acquisition entails the learning and internalization of such rules. Instead, we argue that language acquisition is the by-product of learner-internal (largely linguistic) mechanisms" (Vanpatten & Smith, 2015, p.136).

They stress that the importance of grammatical sensitivity is arguably no longer relevant when we consider L2 acquisition as more than just the learning of new rules. Instead, they postulate that other individual differences play an important role. However, they do not identify any such potential individual differences. In light of these two recent studies, it is clear that the importance of grammatical sensitivity in L2 acquisition has arguably been exaggerated (Vanpatten et al., 2013; Vanpatten & Schmitt, 2015). It should be noted, that both these conclusions are related to L2 acquisition in adults and not secondary school pupils.

3.7.3 Contextual inferences of unknown vocabulary

As noted in section 3.6, one of the metacognitive strategies proposed by Ellis (1995) and Macaro (2003) is the use of contextual cues in order to infer the meaning of unknown words. Trying to infer word meaning from context has been identified as one of the most popular strategies used when encountering new words (Çetinavcı, 2013, p. 2670). One major requirement for this strategy is that the unknown word is noticed during reading, it is not an unconscious process as proposed by Krashen (1982). Contextual inferences is a highly

interesting area of L2 acquisition because it can be thought of as both a process through which vocabulary acquisition occurs, a metacognitive learning strategy, as well as an individual difference; an individual ability unique to every pupil in the L2 classroom (Bolger et al, 2008). Because research has demonstrated that this individual difference can be improved, it is also of great practical value for the L2 teaching profession (Wilson, 2013, p. 30).

It should be noted that the proposed importance of contextual inferences of unknown words has been criticized. Schatz & Baldwin (1986, p. 451) draw several conclusions that criticize the effectiveness of contextual inferences. They assert that "context clues do not reveal the meanings of low-frequency words", and that "context clues appear to be just as likely to result in confusion as in the correct identification of word meaning. Finally, they stress that "context clues work best when the target word is redundant" (Schatz & Baldwin, 1986, p. 451). Another highly important criticism is also made by Macaro (2003, p. 86) who asserts that "[i]nferring the correct meaning of a word in a given context does not necessarily mean that there is retention of the inferred meaning since the immediate communicative need will have been met". In other words, even though it is possible to infer the meaning of unknown words using contextual cues, it does not necessarily mean that the meaning of this unknown word will be remembered. Nevertheless, the importance of contextual inferences has been highlighted by several contemporary researchers (e.g. Bloom, 2000; de Leeuw et al. 2014; Bauman et al. 2003).

According to Bloom (2000, p. 191), the use of contextual cues in order to infer the meaning of unknown words is the most common way of acquiring new vocabulary. He goes far in asserting that most of the words learnt after toddlerhood occurs through this process. In order to illustrate how this metacognitive strategy works, he describes how he himself learnt a new word, *hobbledehoy,* by paying attention to the context in which it occurred. This was not done by noticing the grammatical form of the word (count noun), or by being made aware of this word by someone else. If these were the only cues to the words meaning it could according to Bloom (2000, p. 191) "refer to anything from pickles to phonemes". Instead Bloom (2000, p. 192) assert that "[t]he precise understanding of the word's meaning emerges instead through a sensitivity to the meaning of the passage as a whole. One can tell, for instance, that *hobbledehoy* refers to a kind of person and that it is a bad thing to be".

The use of contextual inferences when reading texts is also an area of interest in cognitive psychology. M. W. Matlin & M. W. C. Matlin (2009, p. 307) propose that this process is important when the reader comes across unknown words. They identify that the context in which the unknown word occurs provides information about when and how it is used (M. W. Matlin & M. W. C., 2009, p. 307). Furthermore, they stress the importance of seeing the word in several contexts. Finally, Bauman et al. (2003, p. 452) argues for the inclusion of instruction into how pupils can use both morphology and contextual cues in order to infer word meaning. However, they also stress that explicit word instruction is necessary.

This assertion is also supported by de Leeuw et al. (2014), who conducted a study exploring the relationship between local and global context, reading tasks, and reader characteristics, in relation to contextual inferences of unknown words. This study is highly influential for this master's thesis as it explores the effect of different learning tasks on the acquisition of vocabulary in relation to individual inferences (vocabulary and working memory). They identified two learner characteristics that were likely to affect learners' contextual inferences. These characteristics are prior knowledge (including vocabulary), and working memory (de Leeuw et al, 2014). The importance of vocabulary in relation to contextual inferences of unknown words is also stressed by Nassaji (2006, p. 396) who in light of the results of his study found that the "results indicate a significant link between depth of vocabulary knowledge and the type and degree of lexical inferences strategy type".

According to de Leeuw et al (2014, p. 276) the acquisition of unfamiliar words rely on contextual cues, and they propose two main forms of contextual inferences: local inferences, and global inferences. The former occurs when the contextual clue either directly precedes or follows the unknown word, whereas the latter occurs when the inference requires that the reader uses knowledge from a larger section of the text (de Leeuw et al, 2014, p.276). Finally, they identify several studies which have shown that local inferences require less time and effort compared to global inferences.

In light of this theoretical framework and previous research, de Leeuw et al (2014, p.279) formulated one research question: "In what way are word learning outcomes influenced by context, task and reader effects, and in what way do these factors interact?" This research question was explored in a pre-test – post-test between subjects research design, where the pupils vocabulary and working memory was tested. The effect of these

individual differences on several different reading comprehension tasks on the understanding of unknown words, relying either on local or global inferences, was then tested. Three hypotheses were formed:

> First, we expected to find words from local contexts to be easier to acquire than words from global contexts (see Cain et al., 2004). Following Kintsch's (2004) levels of text representation, we expected deeper processing tasks to be more successful at enhancing word learning from text than more shallow processing tasks, and all tasks to be better than the single reading of a text. Concerning reader characteristics, we hypothesized that children with more vocabulary knowledge learn more new words than children with less vocabulary knowledge. (de Leeuw et al, 2014, p. 279)

In light of the main finding of their study de Leeuw et al (2014, p. 284) conclude that their first hypothesis was confirmed; the second hypothesis was confirmed in relation to higher processing tasks, but they did not find support for the gap-filling task. The third hypothesis was also partly confirmed. de Leeuw et al (2014, p. 284) assert that "[v]ocabulary knowledge indeed turned out to be important in word learning from text, whereas working memory capacity did not influence the amount of words learnt; not in local contexts and not in global contexts". This was also partly extended to the beneficial use of higher-level tasks for high-vocabulary pupils.

Several aspects of de Leeuw et al.'s (2014) study are highly relevant for this thesis. First and foremost it has shown that individual differences, in this case vocabulary, influences the effect of different learning tasks in relation to contextual inferences of unknown vocabulary. Furthermore, it has shown that local inferences are likely to be more effective than global ones for young language learners. Finally, their concluding thoughts that it "is crucial to help and stimulate children in word learning from text, because vocabulary knowledge lies at the heart of school success" (de Leeuw et al, 2014, p. 285), further cements the importance of vocabulary in any L2 classroom.

3.7.4 The relationship between contextual inferences and other individual differences

In light of de Leeuw et al.'s (2014) research, and the theoretical framework that makes up individual differences in L2 acquisition, I have created Figure 3. It is a graphical illustration of

the possible relationship between other individual differences and pupils' ability to use contextual inferences of unknown words (Bolger et al, 2008; de Leeuw, et al, 2014; Housen et al., 2012; Knight, 1994; Lightbown & Spada, 2013; Robinson, 2002; Skehan, 1989):
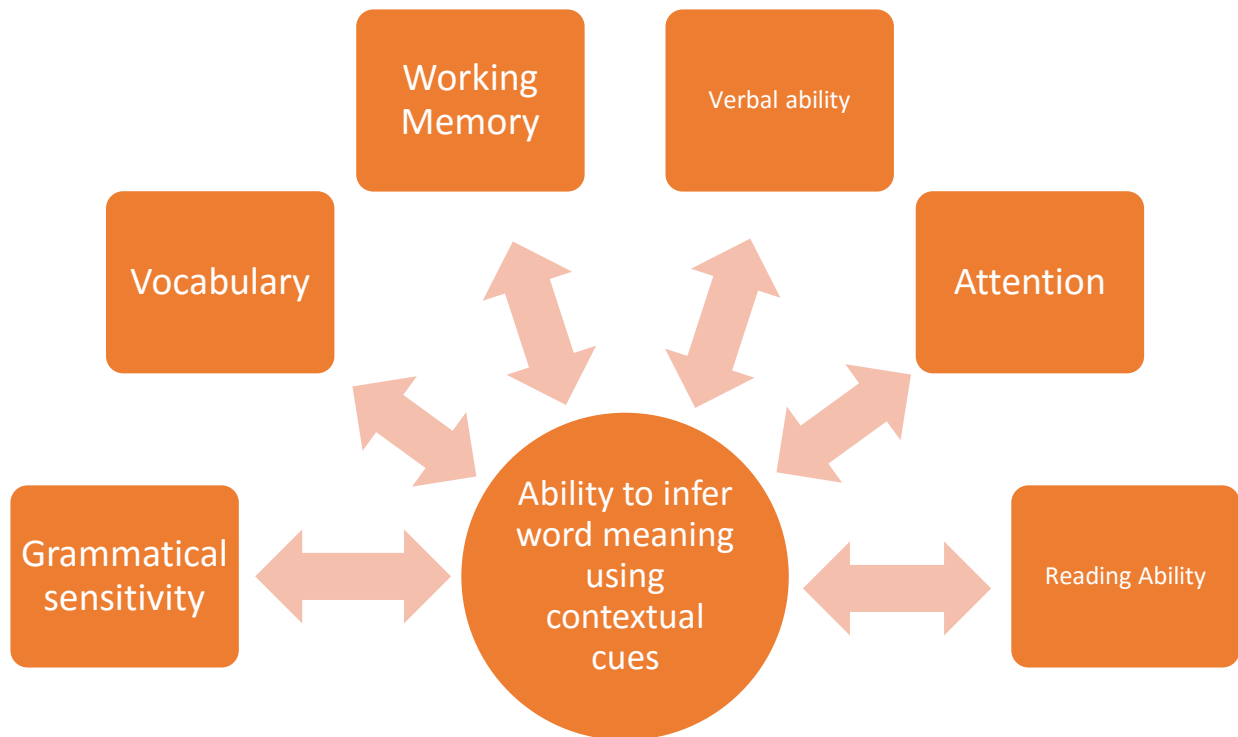
Figure 3 - *Illustration of the possible relationship between individual differences affecting the ability to do contextual inferences of unknown words.*

It is important to note that this figure does not imply a one directional relationship between contextual inferences and other individual differences. Instead, it is for example reasonable to assume that although vocabulary size has been shown to affect contextual inferences of unknown words, it is also possible that being adept at using contextual cues in order to infer word meaning might cause a higher vocabulary size. The inclusion of working memory might perhaps be controversial in light of de Leeuw et al.'s (2014) conclusions. However, it is not possible to rule out that it might be a potential component under different circumstances. Finally, verbal ability has been included in light of Knight's (1994) research into the relationship between verbal ability, dictionary use, and vocabulary item acquisition.

3.8 Theory – summary

This chapter has presented the theoretical background and previous research which has formed the foundation of this investigation. The first half of this chapter has presented the main terms used in vocabulary research, different levels of vocabulary knowledge, the

45

relationship between vocabulary size and reading comprehension, and how vocabulary is acquired. Additionally, this section has presented contemporary research into the usefulness of using explicit focus on vocabulary.

The second half of this chapter has been oriented towards individual differences between pupils in an L2 setting. It has presented a number of different individual differences which can potentially play a role in how well a second language is acquired. Individual differences related to motivation, language proficiency, language aptitude, and ability to use contextual cues to infer word meaning from context have been presented in detail. Finally, contemporary research into different areas of contextual inferences has been identified and presented.

How the results of this investigation relate to this theoretical framework is explored in section 6.2. Finally, how this theoretical framework can be incorporated into new potential areas of research can be found in section 6.4.

CHAPTER 4 – METHOD

This chapter will present the method employed in order to realize the aims of this master's thesis. This includes the testing instruments, as well as the teaching plans and homework for the two group conditions. The selection criteria for the groups and teacher that were part of this research project, as well as the selection criteria for the text about global warming and climate change (National Geographic, 2015) will also be presented. Finally, the measures employed by the researcher to ensure an adequate level of validity and reliability will be identified.

4.1 The quasi-experiment

The research questions explored in order to achieve the aims of this study are all linked to the relationship between different variables, and how they affect each other. Thus, it is necessary to use a quantitative approach. In light of this, the most suitable approach in order to answer these research questions is to design and conduct an experiment or quasi-experiment. As the main justification for conducting this study is to provide information that can be of practical value for practicing teachers, I decided to conduct the research in a naturalistic setting, in a Norwegian lower-secondary school. This meant that it would not be possible to use random group assignment, because this would interfere with the pupils' learning environment. Such a change in the learning environment could possibly become a confounding variable[10]. It could also be considered too great an interference with the pupils' school time. This could make them less likely to participate. Thus, it was not be possible to conduct a full-scale experiment. Instead, a pre-test, post-test, delayed post-test quasi experiment using two group conditions was designed.

It is important to keep in mind that there is a limited number of teachers available in the Norwegian educational system who are willing to donate their limited and valuable teaching time for research purposes. In light of this, researchers who want to conduct research in a naturalistic setting cannot be too picky in terms of the selection criteria for the teachers. Even so, some selection criteria are necessary in order to stay within the aims and scope of

---

[10] A confounding variable is defined by Dancey & Reidy (2014, p. 10) as "a specific type of extraneous variable that is related to both variables that we are interested in".

the research that they want to conduct. However, because of the limited availability of willing teachers there is also an incentive to collect as much usable data as possible during classroom research. Thus, I have tried to design my quasi-experiment in such a way that it collects the data in a manner which allows for a considerable number of variables to be explored. The variables to be explored in the quasi-experiment are: (1) Group condition, consisting of teacher-initiated explicit focus on vocabulary (experiment group) and focus on content (control group); (2) vocabulary size; (3) ability to infer word meaning from context (individual difference); (4) pupil understanding of text; and (5) acquisition/learning of new vocabulary. Figure 4 is a graphic representation of the research design.
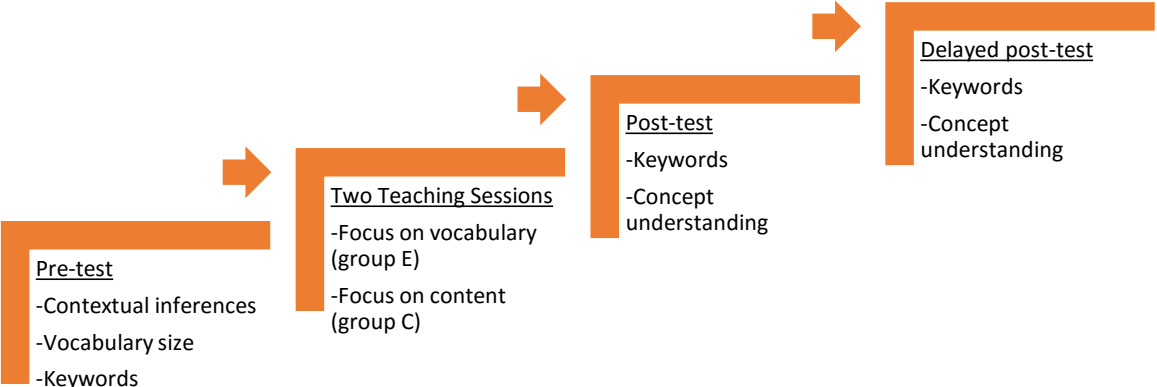
Figure 4 - *Illustration of the Quasi-experiment design.*

The four research questions were investigated by means of null hypothesis testing using different statistical analyses. The following analyses were computed: t-test, correlation, and linear regression modeling. Q1 and Q2 were explored using independent-means t-tests in order to measure the significance of potential group differences as concerns the variables comprising the post-test (i.e. understanding of text, and learning of keywords). Q-3 and Q-4 were investigated using Pearson's correlation and simple linear regression modelling, respectively (consult section 1.2 for an explanation of the research questions).

4.2 Teacher selection criteria and sample size

Because of the decision to conduct the quasi-experiment in a naturalistic setting, it is very important to identify and avoid as many confounding variables as possible. This must be done in order to ensure that any potential group differences on the post-test are a result of the binary variable (explicit focus on vocabulary vs focus content). Additionally, the selection

criteria for the teacher and groups are crucial. The decision to have one teacher teaching two groups, as opposed to two teachers teaching one group each, was made in order to avoid the teacher as a confounding variable. If two teachers had been utilized there is a chance that potential group differences could be attributed to different teaching styles. This could potentially confound the results.

The groups that were part of the quasi-experiment are pre-defined, and dependent on the teacher. Therefore, the selection criteria for the teacher also determined the selection criteria for the groups. The selection criteria for the teacher were that (1) the must teach English in a Norwegian lower secondary school (8-10th grade). (2) The teacher must teach English in at least two classes at the same year level. (3) In order to ensure an adequate sample size in the two group conditions, the classes taught by the teacher must contain at least 25 individual pupils. Other factors related to the teacher such as age, gender, number of years of teaching experience, and teaching competence in other subjects were considered. However, they were deemed to be irrelevant and cut in order to avoid recording any personal information, or information that can be used to triangulate the names of the teacher or pupils in the quasi-experiment. It should be noted that although the preferred sample size in each group is at least 30, the typical Norwegian classroom rarely contains more than 30 pupils. Thus, there is a chance that the small sample size will affect the result of the statistical analyses and their generalizability. This will be addressed further in sections 6.1 and 6.4.

The teacher selected for the quasi-experiment in this thesis is an experienced teacher at a Norwegian lower-secondary school, who teaches several classes at the same year level. Two of the teacher's classes at the same year level were selected to function as the experiment and control group. These classes each contained 30 individual pupils. Informed parental consent was collected in both group conditions, using a consent form sent as a letter home three weeks in advance of the data collection. Those who did not wish to participate, or failed to reply to the consent form, were not part of the testing.

4. 3 Text selection criteria

The major teaching material used during the teaching in the two group conditions is the text about global warming and climate change. This text formed the basis for the two teaching sessions, and the selection criteria for this text are therefore crucial. The selection criteria were as follows: (1) the text used as teaching material in the two group conditions must be

related to the topic global warming and climate change. (2) The text should be available online in open access. (3) The text used as teaching material must contain authentic written English language. (4) The text should contain no less than 500 and no more than 900 tokens[11].

Preferably, the text would also contain vocabulary which is close to or slightly below the 95-98% mark identified by Laufer (1989), Lightbown & Spada (2013) and Macaro (2003). This would make the text challenging but not incomprehensible for the pupils. Additionally this would mean that roughly 5% of the text would contain words which are unknown to the pupils. These lexical items could be learnt either explicitly through teacher-initiated keyword identifications, or implicitly by, for example, inferring from context. However, since the vocabulary size of the pupils is unknown before the pre-test, finding a text containing roughly 98% known vocabulary was not be possible.

The text that was finally chosen for the teaching in the two group conditions was adapted from *National Geographic Education* website (National Geographic, 2015), and can be found in Appendix 3. It is an encyclopedic entry, which introduces the concepts of global warming, the greenhouse effect and effects of global warming, future changes, and what we can do. In order to make the text conform to the selection criteria parts of the text had to be removed. The two final sections were removed from the text, as well as some paragraphs. The shortened text contained 849 tokens. Two illustrations were also added, one showing a sweating earth, and one detailing the processes in the greenhouse effect.

This text was chosen because it conformed to all of the selection criteria (after some editing). Furthermore, it presented the concepts of global warming and climate change in a structured and logical manner, and was written by a well-respected organization. Interestingly, the text also features a list of vocabulary items that had been identified by the authors as potentially difficult, or unknown for younger readers. Some of these vocabulary item were identified by me as keywords, and were therefore taught in the experiment group condition[12].

---

[11] See section 3.2 for an explanation of tokens.
[12] Potentially difficult words that are likely necessary to understand in order to fully benefit a text and its content.

50

4. 4 Pre-test measurement instruments

The pre-test (Appendix 4) was conducted over the course of one English session (45 minutes), and was divided into three components. The three components were (1) vocabulary size, (2) ability to infer word meaning from context, and (3) previous knowledge of keywords. They will serve three purposes: (1) ensuring group condition similarity, (2) provide two independent variables (the individual pupil's vocabulary size, and ability to infer word meaning from context), and (3) answering Q-3 and Q-4 using these two variables. The use of intact groups means that there is a potential for the two groups to be different on some unknown but critical factor. Three such possible factors were identified and measured during the pre-test. These were vocabulary size, knowledge of keywords, and ability to infer word meaning from context. A null hypothesis was formed postulating that the two groups will not differ in their vocabulary size, ability to infer word meaning from context, and previous knowledge of keywords. Three independent means t-tests were then used in order to accept or reject this null hypothesis. The results of these analyses can be found in section 5.2.

Arguably, it could also be interesting to measure pupils' reading comprehension. Because one component of the post-test measures pupils' understanding of the content of the text on global warming and climate change, this could be thought of as a possible group difference that could confound the results. However, there was a limited amount of time available to administer the pre-test. Adding a fourth component would give the pupils very little time to finish all the sections, and could affect the results. Furthermore, reading comprehension tests are often very detail oriented, at a sentence level. Nevertheless, it should be kept in mind that the two groups can differ in their reading comprehension skills even though this has not been measured.

4.4.1 Pre-test measurement instruments – Vocabulary size

The pupils' vocabulary size was measured using a standardized receptive vocabulary size test (VST). The VST was developed by Nation & Beglar (2007). Several other tests were considered, notably the vocabulary levels test (VLT) (Schmitt, 2010, p. 279) and two alternative VST tests developed by Nation (2012). The VST test was finally chosen because it would arguably lead to a more precise measure of the pupils' vocabulary. It provides an estimate of pupils' vocabulary ranging between the 1000 and 20 000 most common word families. In contrast, the VLT provides a measurement of the pupils' vocabulary at different levels (less than 2000,

5000 and 10 000). As this inquiry is focusing on individual differences, one can argue that the VST will provide a greater degree of nuances between the pupils.

Nation (2012) provides three alternative monolingual VSTs, one measuring vocabulary size up the 14 000 word level, and two measuring vocabulary size up to the 20 000 word level. Because of time constraints on the pre-test, the 20 000 word level test was used. It is according to Nation (2012) appropriate for all proficiency and year levels. This test does not distinguish between the four commonly used word lists (high/low frequency, academic, and technical), or measure the pupils productive, or oral vocabulary. Instead, it provides a measure of pupils' receptive vocabulary size by using a multiple choice test with 100 questions. The number of correct answers is multiplied by 200 in order to provide an estimate of the pupils' vocabulary size.

4. 4. 2 Pre-test measurement instruments – Ability to infer word meaning from context

Because of the limited time frame and scope of this particular thesis, there will be no time to conduct a pilot study. In light of this, one major concern during the design phase has been to find standardized tests to use in the pre-test. This has been challenging when it comes to the instrument measuring pupils' ability to infer word meaning from context. Eventually, it was decided to use an unknown word in sentences test (Appendix 4). This test was developed from a test used by Wilson (2013, p.35) in her master's thesis. There was no contact information found in her thesis, so it has not been possible to contact her to gain permission to use her test. Her test was developed using an internet teacher resource (Help Teaching, 2013), and contains 8 words in sentences which the pupils need to guess the meaning of by means of contextual cues.

The test was developed for an American 3rd grade, but the words used are arguably unfamiliar to, or used in a way that is uncommon for, Norwegian lower secondary school pupils. It was chosen because it provided a viable measure of pupils' ability to infer word meaning from context in Wilson's (2013) thesis, and showed that the pupils' had improved this ability after a period of teaching focusing on how to use contextual cues (Wilson, 2013). Because of this, it is likely that the test provides an adequate measure of pupils' ability to infer word meaning from context.

52

Some changes have been made to this test in order for it to be relevant to this thesis. The first change was to remove the component of the test where the pupils should identify what contextual cues they used to infer the word meaning. Another change is the addition of three new sentences using constructed non-words (*blayned, peltively, sconizatic*). These words are shown in context and follow the English rules for word construction. The forms of the words are either verb, adverb, or adjective (Hannisdal & Jenset, 2012). They were added in order to provide some words that were absolutely certain to be unknown to the pupils, regardless of their vocabulary. The constructed words and sentences were collected from the book *Analyzing English Grammar – Exercises for Advanced Students* (Hannisdal & Jenset, 2012). Because the constructed words do not exist, the correct answers for them are problematic. Nevertheless, in light of the context in which they are presented, it is fairly clear that their meaning and function is close to *painted, carefully*, and *angry/surprised*. Word meaning explanations that are close to these will be considered correct answers.

Because the unknown words are presented in context using only one or two sentences, it is likely that the pupils will have to rely on local rather than global inferences as presented in section 3.6. As local inferences have been shown to require less time and effort compared to global inferences (de Leeuw et al., 2014), this will likely be less time consuming and demanding for the pupils. In light of the limited time available for the pre-test, local inferences are most likely more suitable compared to global inferences. However, it should be noted that the test used in this study has not measured pupils' ability to make contextual inferences based on global cues.

The scoring procedures for this test are inspired by those used by Carpenter, Sachs, Martin, Schmitt and Looft (2012, p. 83). Two full points were awarded for answers that corresponded with the meaning of the target word (being synonyms or near synonyms), and caused a correct understanding of the overall sentence. One point was awarded to answers which did not correspond exactly to the target word without causing the overall meaning of the sentence to be incorrect. Finally, zero points were awarded if the response word did not have the correct syntactic function and/or differed so much in meaning from the target word that it would lead to a misunderstanding of the sentence.

In order to illustrate the scoring system, one sentence in the words in sentences test is *Thomas hiked to the APEX of the mountain and got a great view of the entire city*. A correct

53

answer to this sentence, resulting in a full two points, would be a synonym or near synonym such as *top, peak, summit etc.* A near correct answer could for example be *viewpoint* or *lookout,* and would result in one point being scored. Finally, a wrong answer would be any word which is not a synonym or near synonym of the target word, and whose meaning would cause a wrong understanding of the sentence, such as *bottom or cave.*

One important aspect of this scoring system is that it does not distinguish between degrees of wrongness. Thus, even though *bottom* or *cave* can be thought of as more related to words such as *hiking* and *mountain*, they would be scored similarly to words such as *scuba diver*, *wristwatch*, or *green*. Similarly, the word's grammatical function is not taken into consideration in this scoring system.

4. 4. 3 Pre-test measurement instruments – Previous knowledge of keywords

The third component of the pre-test was added in order to be certain that the two groups had more or less similar previous knowledge of the keywords. Thus, any learning of these keywords would be a result of them being seen in context when the pupils read the text on global warming and climate change, or taught as a keyword. This section uses a multiple choice test with word definitions and no context, in order to measure any previous knowledge of the keywords. It is important to keep in mind that this means that the pupils have been exposed to the keywords before the teaching procedures start. However, they will not be provided with any definitions of the words, or see them in context. Thus, it is believed that this will most likely not lead to any incidental learning before the teaching procedures start. Furthermore, both group conditions will be exposed to the keywords in the pre-test, and any potential group differences on the post-test will therefore not be a result of this.

The keywords test consists of 12 words. These words were shown without context. A multiple choice design was used, with one correct answer and three distractors. The answers are made in the form of definitions, and the pupil must chose the correct definition for each keyword. The distractors were made using definitions of words belonging to the same word class as the correct answer. They were either retrieved from the online oxford advanced learners dictionary, or made by the researcher.

4.5 Post-test measurement instruments

The post-test used in this quasi-experiment was designed to measure two areas of the pupils' learning. These areas were learning of keywords, and understanding of the main concepts in the text about global warming and climate change. The pupils' understanding of keywords was measured using the same test as in the pre-test. However, the order in which the questions were asked was randomized. This was done in order to avoid pupils recognizing the questions from the pre-test.

The pupils' understanding of the main concepts found in the text was also measured using a multiple choice test. Once again questions were asked using one correct response and three distractors. The distractors were formed in such a way that they were similar to the correct response in length, and used the same vocabulary and focus as the correct answer. The questions were based on the text about global warming and climate change used during the two teaching sessions. Each question was roughly centered on one paragraph, and the correct answer was formulated without directly copying it from the text. The questions were also asked in a randomized order as compared to how they were presented in the text. This was done in order to avoid pupils simply answering questions that they recognized as being from the text. Finally, one productive question was asked. This question was related to the concept of greenhouse gasses, and asked the pupils to explain how the earth's atmosphere was similar to the glass walls in a greenhouse. However, this question was not used for the analyses in this investigation.

4.6 Teaching procedures

Because this quasi-experiment uses an intact group design with two group conditions, it was necessary to develop two sets of teaching procedures. The major difference between these two teaching procedures was the methods and teaching activities used in the two group conditions. In group E, the teaching was based on the presentation and teaching of keywords, using a combination of definitions, examples from the text, and illustrations. The teaching procedures in the control group focused on the major concepts that were presented in the text, and no keywords were taught or worked with. Teaching plans detailing the teaching activities for both group conditions were presented to the teacher. These can be found in Appendix 5.

I produced two different sets of PowerPoint presentations. These were utilized by the teacher in the two group conditions. Since the teaching in the two groups was conducted by a teacher, and not by the researcher, it was important to ensure that the teacher-led presentations were conducted in such a way that they were in accordance with the teaching plans and group conditions. By using PowerPoints prepared by the researcher, it was possible to provide a framework for the teaching presentations that were in compliance with the two different group conditions.

4.6.1 Teaching procedures – experiment group

The main teaching procedures used in the experiment group were focused on the presentation, identification, and use of keywords. In the first session, the teacher held a 20-minute PowerPoint presentation, where the goals and topic of the session were presented. After a brief introduction to the concept of global warming and climate change, the presentation moved on to a list of the keywords found in the text. Six of these were presented using illustrations, definitions, and sentences from the text (*Global, climate, atmosphere, greenhouse effect, emission, and sea level*). Concepts and sentences from the text were used in order to explain these keywords. After the presentation, a list of all the keywords was left on the smartboard. The pupils were then instructed to read through the text about global warming, and underline any keywords that they found in the text, as well as any other unknown words. At the end of the session the pupils were informed of the homework, and asked to write it down. Their homework was to read through the text at home, and then construct and write down their own sentences using the keywords.

The first part of the second teaching session was, once again, led by the teacher using a PowerPoint presentation. The teacher first started by asking the pupils to read out some of the sentences that they had made as part of their homework. After this, the list of keywords was repeated once more. The remainder of the PowerPoint presentation was used to further teach three more keywords using illustrations, definition, and sentences from the text (*agricultural, fossil fuels, glacier*). After this, the pupils were instructed to complete a "fill in" exercise that was handed out (Appendix 6). This fill in exercise consisted of sentences from the text about global warming, where one or two keywords had been removed. The pupils' task was to fill in these missing words using the text and the keywords sheet. When this exercise was finished, they were told to construct their own factual sentences about global

warming and climate change, using the list of keywords and the text as inspiration. This was also their homework.

4.6.2 Teaching procedures control/content group

Because the control group did not receive any teacher-initiated focus on vocabulary, considerably more teaching time was spent on presenting the main concepts found in the text. As in the experiment group, all of the teacher-led activities were conducted using PowerPoint presentations. Additionally, the two teaching sessions were organized in the same manner as in the experiment group. This meant that the length of the PowerPoint presentations were very similar to the experiment group. Finally, the slides used in the experiment group condition which did not contain explicit focus on vocabulary were also used in the control group. Thus, any differences between the teaching procedures in the two group conditions are related to the explicit focus on vocabulary versus increased focus on content.

During the first session, four concepts from the text on global warming were presented using illustrations and partner discussion. These concepts were *global warming, greenhouse effect, causes of global warming,* and *causes of greenhouse gasses*. Although there was no explicit focus on the keywords in the control group, they were still used during the presentations of the key concepts. Thus, any learning of these keywords are most likely the result of incidental learning, or contextual inferences either during the presentations, or through reading the text about global warming. After the PowerPoint presentation, the global warming text was handed out and the pupils were tasked with answering three questions. These questions were (1) *what is global warming,* (2) *what is the greenhouse effect,* and (3) *how can humans cause the earth's temperature to rise?* Those pupils who were not finished after the session were told to do it as their homework.

The second session in the control group followed a similar pattern to the one in the experiment group. The first part of the presentation focused on repeating the four main concepts that were presented during the first session. After this, three more concepts from the text were highlighted (*how can humans cause global warming, what are the greenhouse gasses, what are the effects of global warming*). The pupils were then instructed to work with a "true or false" task prepared by the researcher (Appendix 7). This true or false task consisted of 10 statements about global warming and climate change. In order to assert whether or not

57

these statements were true, the pupils had to use the text about global warming and climate change. Finally, the pupils were instructed to summarize the text in their own words.

4.6.3 Teaching procedures – Summary

This section has presented the differences between the two group conditions. Any group differences in the post-test results are expected to most likely be a result of these differences. A summary of the key-differences can be found in Table 1.

Table 1- Summary of the teaching procedures in both groups.

Experiment Group	Control Group
Teacher-led presentation of keywords	Teacher-led presentations of concepts
Identify the keywords in text task	Answer conceptual questions using text task
Create own sentences using keywords task	---
Fill in missing keywords task	True or false statement task
Construct factual sentences using keywords task	Write summary of text task

4.7 Validity and reliability measures

Loewen & Philp (2012, section two) stress the importance of not only focusing on the end results of a study, but also the processes that have caused these results. They assert that the use of interventionist quasi-experiments allow researchers not only to measure the results of two different classroom practices, but also to manipulate the methods and activities used. In order to minimize the likelihood of any unknown pre-existing group differences affecting the final results, it is therefore necessary to identify and address any potential confounding variables which might interfere with the validity and reliability of the findings of the study. This section will present and address four variables that have been identified as potentially confounding variables. These variables are group differences, classroom noise level, deviation from teaching schedule, and deviation from teaching plan.

4.7.1 Group Differences

Because this research project was conducted in a classroom setting, there was no possibility of randomly assigning the participants into the two group conditions. Because intact groups were used, it is not possible to assume that the groups are equivalent (Brown & Rodgers, 2002, p. 212). However, as the two groups in this study are from the same school, at the same year level, and having the same English teacher it is likely that they are not too different. As there

is a limited amount of time available for conducting the study, it is not possible to test for every potential difference between the groups.

Because this study focuses on vocabulary and individual differences, I decided to measure three variables on the pre-test. These variables were contextual inferences, previous knowledge of keywords, and vocabulary size. By measuring these variables on the pre-test it is possible to address any potential group differences related to them. Nevertheless, it is still possible that some form of unmeasured, but critical, difference between the groups already exist. Unfortunately, this is one of the major limitations of a quasi-experiment design. Even so, because there is need for research into instructed second language acquisition in a natural setting, this approach is arguably necessary.

4.7.2 Classroom noise level

One potential variable which is not possible to measure on a pre-test is the level of classroom noise during the teaching procedures in the two group conditions. If there is consistently more noise in one of the two group conditions, this might affect the post-test results. In order to measure any potential group differences on this variable, an observation schedule measuring the noise level on a scale from 1-5 (low to high) every 10 minutes was created and used by the researcher during the two teaching sessions in each group. Additionally, it was possible for the researcher to write detailed field notes during the sessions.

The measured noise level in the two groups during the teaching sessions never exceeded *2* which indicates a low level of classroom noise. The observation schedule can be found in Appendix 8. In light of these observations, it is unlikely that any potential group differences in the post-test is a result of differences in classroom noise levels.

4.7.3 Deviation from schedule

Because the teaching in the two group conditions was conducted by their English teacher and not the researcher, deviations from the time schedule as designed by the researcher were a possibility. In order to avoid this, the teaching plans and PowerPoints were reviewed and explained to the teacher by the researcher. Additionally, an observation schedule focusing on any potential deviations from the time schedule was designed and used during the two teaching sessions in the different group conditions.

59

By using running observations, several deviations from the time schedule were discovered. During the first sessions, it was discovered that the PowerPoint presentations in both groups each took 20 minutes longer than planned. However, this happened in both group conditions and should therefore not affect the post-test results. The second session in the experiment group was divided into two 30-minute sessions. Additionally, several pupils in the control group were late for the second teaching sessions due to icy road conditions. Nevertheless, these deviations from the time schedule are likely to be within what should be expected in a classroom setting. In light of this, it is doubtful that they have affected the results in any significant way. Still, they should be noted as a possible confounding variable.

4.7.4 Deviation from teaching instructions and plans

The final confounding variable identified by the researcher was deviations from the teaching instructions and plans. If the teaching plans, activities, and instructions created for the two groups conditions were not followed correctly, this could be a major problem as concerns the validity of the results in this study. In order to avoid such deviations, I provided the teacher with detailed teaching plans, pre-made teaching activities, and PowerPoint presentations. Furthermore, an observation schedule and debrief was made in order to provide the researcher with both an emic and etic view of how the teaching procedures in the two groups were conducted. By debriefing (Appendix 9), it was also possible to gain insight into the teacher's experience of using the two teaching procedures. However, because of the limited space and scope of this thesis, it is not possible to pay much attention to this.

It is important to note that teachers are not programmable robots, and will probably always impart some of their own teaching practices unrelated to the instructions for the two group conditions. Thus, it is not possible to expect them to follow the teaching instructions and plans flawlessly. This in itself is not likely to affect the validity of the results in this study. Even so, any deviations from the teaching plans that directly contradict the group conditions, with particular emphasis on any explicit focus on vocabulary in the control group and on content in the experiment group has been noted. The deviations from the teaching instructions and plans can be found in table 2:

Table 2 - Deviations from teaching plans.

Session	Group Condition E	Groups Condition C
1	-Asks the pupils to read out loud. -Uses mobile phone to find out size of Jostedalsbreen.	
2	-Spends time detailing oil production.	-Focuses on word *oxygen* -Asks the pupils to create sentence using *emissions.*

In light of the observations, there has not been many deviations from the teaching plans. The focus on the size of Jostedalsbreen, and the reading out loud activity might have taken up some of the time in the experiment group, but not to such an extent that it should affect the results. The keyword *emissions* was part of the post-test. It should therefore be noted that some time was used focusing on this word in the control group, and this might have an impact on the results. Nevertheless, this is only one of the 12 keywords that were tested in the post-test.

CHAPTER 5 – DATA ANALYSIS AND INTERPRETATION

This chapter will present the findings of this research project. It will focus on the data that has been collected during the quasi-experiment which forms the foundation of this study. The chapter will be divided into two sections. The first section will focus on descriptive statistics, providing an insight into the confidence intervals and data distribution in the pre-test, post-test, and delayed post-tests. The second section will focus on the inferential statistics used to answer the four research questions that are explored in this master's thesis. All statistical tests were computed using IBM SPSS version 23.

5.1 Descriptive statistics

This section will serve two purposes: to give a representation of the scores on the tests conducted during this inquiry, and present the distribution of scores in the two groups. In order to conduct the inferential statistics used to answer the research questions, it is necessary to demonstrate the distribution of the data collected. Because the sample size in both groups is less than 50, the most appropriate test of normality is the Shapiro-Wilks test (Dancey & Reidy, 2014). By testing the hypothesis that the sample is normally distributed, this test assesses the normality distribution of the sample (Shapiro & Wilks, 1965). Thus, if the Shapiro & Wilks test results in a *p* value below 0.05 the hypothesis is rejected, and the sample is most likely *not* normally distributed. In other words, a *p* > 0.05 indicates that the sample is normally distributed, and parametric analyses are possible. However, it is also necessary to examine the normal Q-Q plots in addition to the Shapiro-Wilks test of normality.

5.1.1 Data Distribution Pre-test – Contextual inferences of unknown words

The first part of the pre-test assessed pupils' ability to infer the meaning of unknown words using contextual cues. The maximum score of the test was 18 points. In group C, the contextual inferences pre-test mean score was 8.5, with a standard error (*SE*) of .786[13] and a 95% confidence interval of 6.85–10.15.[14] This indicates that it is possible to generalize the mean score of this sample to the population as being between 6.85 and 10.15 with a 95% degree of

---

[13] Standard error is according to Dancey & Reidy (2014, p. 119) calculated by "dividing the standard deviation of the sample with the square root of the sample size", and is used to calculate the 95% confidence intervals.
[14] The 95 % confidence interval provides a measurement of the likely mean score of the population.

certainty. The standard deviation (*SD*)[15] was 3.51, and the minimum score was 1 and the maximum score was 16 (range 15)[16]. According to Dancey & Reidy (2014) 70 % of the sample are likely to fall within one standard deviation. Thus, in this sample 70 % of the scores are likely to be between 8.5 +/- 3.51.

Figure 5 shows the score of the contextual inferences section of the pre-test in group C. The X-axis shows the ID[17] of a single participant, and the Y-axis shows the score of the contextual inferences test.
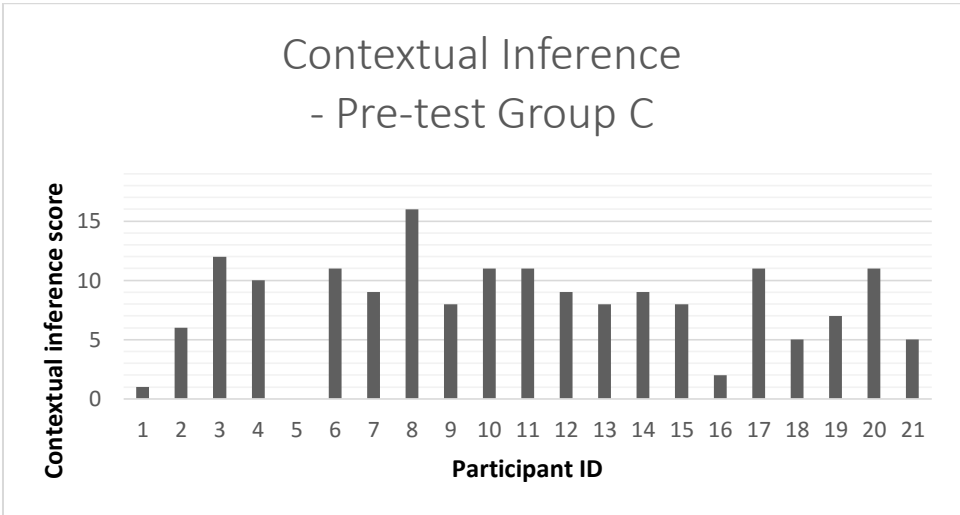
Figure 5 - *Bar chart illustrating the contextual inference scores in group C.*

In group E, the contextual inferences pre-test mean score was 7.72 (*SD =3.53, SE = .706)* with a 95% confidence interval of 6.26-9.18. The minimum score was 0 and the maximum score was 13 (range = 13). Figure 6 shows the scores of the contextual inferences section of the pre-test in group E. The *X-axis* shows the participant ID, and the Y-axis shows the contextual inference score.

---

[15] " *Standard deviation* is defined by Dancey & Reidy (2014, p. 79) as "the degree to which the scores in a dataset deviate around the mean. It is an estimate of the average deviation of the scores from the mean".
[16] *Range* is defined by Dancey & Reidy (2014, p. 77) as "the difference between the minimum and maximum scores.
[17] For the purpose of all graphical representations of data, ID should be understood as identification and not individual difference unless stated otherwise.
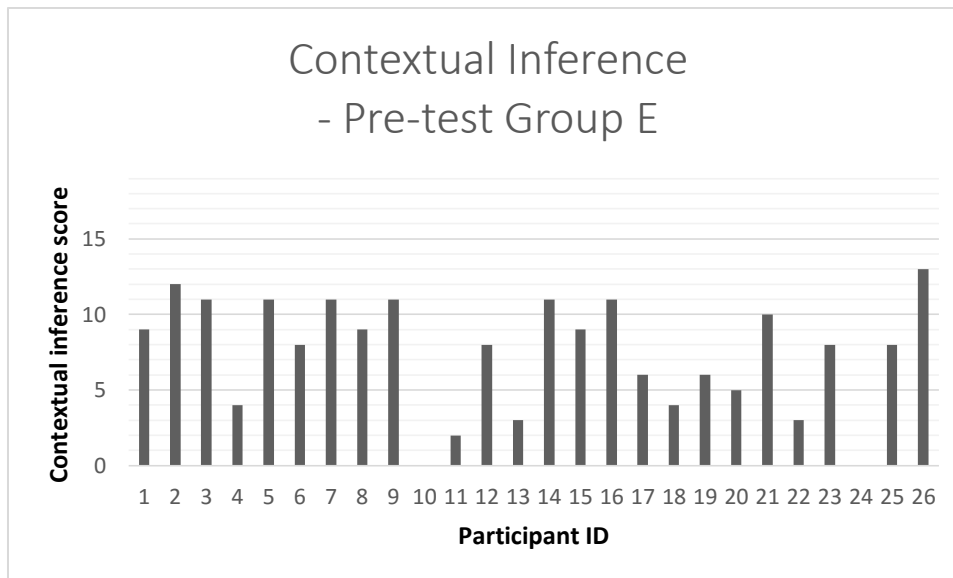
Figure 6 - *Bar chart illustrating the contextual inference scores in group E.*

In order to assess the normality of the score distribution in both group conditions, a Shapiro-Wilk test of normality was conducted. The results of this test is summarized in table 3:

Table 3 - Shapiro-Wilks test of normality on the contextual Inferences component of the pre-test.

Group	Shapiro-Wilk		
	Statistic	*df*	Sig.
C	.953	20	.422
E	.934	25	.107

In the Shapiro-Wilks test of normality, degrees of freedom (*df*) is the sample size -1, and the significance (Sig.) indicates to which extent the formal hypothesis can be rejected. In light of this analysis $p > 0.05$ (.422 and .107) and the formal hypothesis is not rejected. Thus, the data can be considered normally distributed.

5.1.2 Descriptive statistics – Keywords tests

In order to assess the pupils' existing knowledge of the keywords, a 12 point multiple choice test was administered as part of the pre-test. In group C, the mean score of the keywords pre-test was 6.55 (*SD = 2.37, SE = .544)* with a 95% confidence interval of 5.30–7.59. The minimum score was 3 and the maximum score was 11 (range = 8). In the keywords post-test, the mean score increased to 7.56 (*SD = 2.33, SE = .550)* with a 95% confidence interval of 6.40–8.72. The minimum score was 3 and the maximum score was 10 (range 7). Finally, in the delayed post-

test the mean score increased to 7.90 (*SD = 1.97, SE = .441)* with a 95% confidence interval of 6.98–8.82. The minimum score was 4 and maximum score was 11 (range = 7).

Figure 7 shows the development of the pupils understanding of the keywords throughout the quasi-experiment in group C. Each bar represents the keywords score on the pre-test, post-test, and delayed post-test, respectively. The X-axis represent the participant ID, and the Y-axis shows the keywords score.
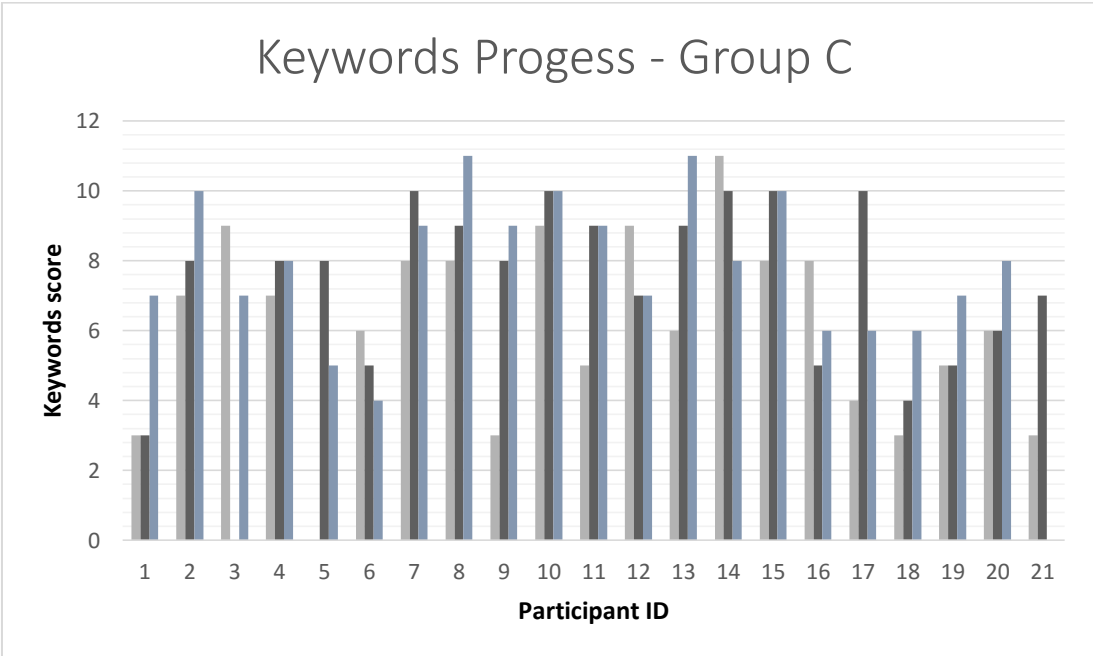
Figure 7 - *Bar chart illustrating the keyword score progress (pre-test light grey, post-test dark grey, delayed post-test blue) in group C.*

In group E, the keyword mean score on the pre-test was 4.61 (*SD = 2.31, SE = .544)* with a 95% confidence interval of 3.45–5.77. The minimum score was 0 and the maximum score was 8 (range = 8). In the post-test, the mean score increased to 7.91 (*SD = 2.64, SE = .551)* with a 95% confidence interval of 6.77–9.06. The minimum score was 2 and the maximum score was 11 (range = 9). Finally, in the delayed post-test, the mean score increased to 8.15 (*SD = 2.37, SE .456)* with a 95% confidence interval of 7.21–9.09. The minimum score was 4 and the maximum score was 12 (range = 8).

Figure 8 shows the development of the pupils' understanding of the keywords throughout the quasi-experiment in group E. Each bar represents the keywords score in the pre-test, post-test, and delayed post-test, respectively. The X-axis represent the participant ID, and the Y-axis shows the keywords score.
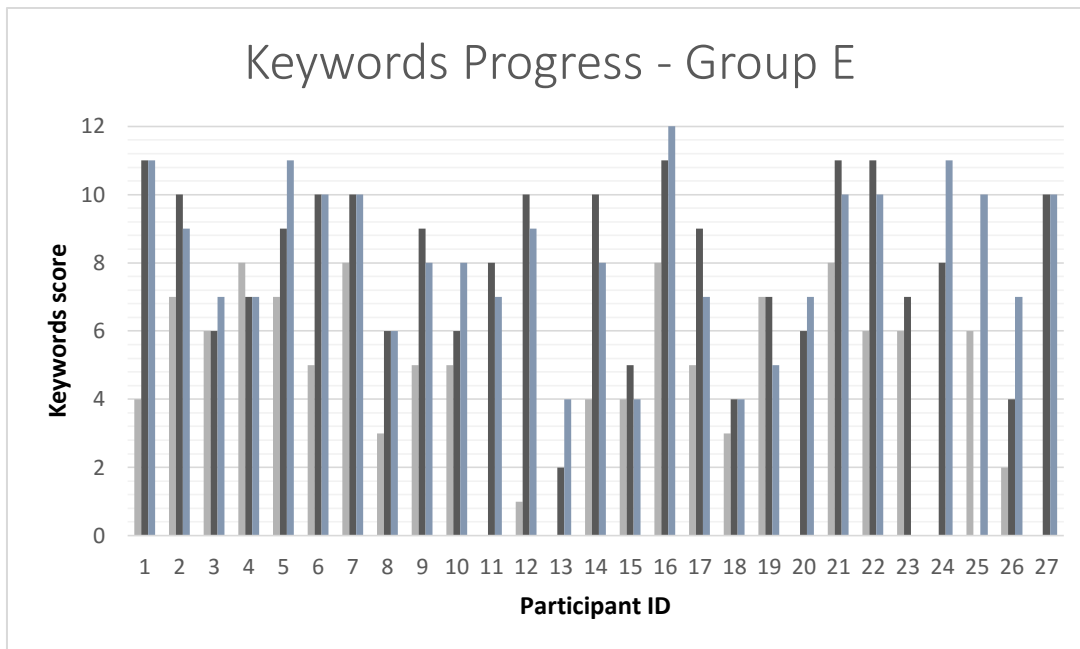
Figure 8 - *Bar chart illustrating keywords score development (pre-test light grey, post-test dark grey, delayed post-test blue) in group E.*

Because the keywords tests will be used for several inferential statistical tests, it is necessary to assess the distribution of the scores in both groups. In order to assess the normality of the score distribution in both groups, a Shapiro-Wilk test of normality was conducted. The results of this test is summarized in table 4.

Table 4 -Shapiro Wilk test of normality on the keywords tests.

Test	Group Condition	Shapiro-Wilk		
		Statistic	*df*	Sig.
Pre-test Keywords	C	.950	18	.428
Pre-test Keywords	E	.918	23	.059
Post-test Keywords	C	.882	18	.029
Post-test Keywords	E	.911	23	.043
Delayed post-test Keywords	C	.957	18	.537
Delayed post-test Keywords	E	.947	23	.258

In light of these data, the formal normality hypothesis is rejected for both group conditions in the post-test. However, this is only at the .05 level, and for group condition E this is barely below this threshold. Thus, in order to assess if these distributions are not normalized, it is

necessary to examine the normal Q-Q plots for both groups[18]. Figure 9 shows the Q-Q plots for both groups on the post-test keywords.
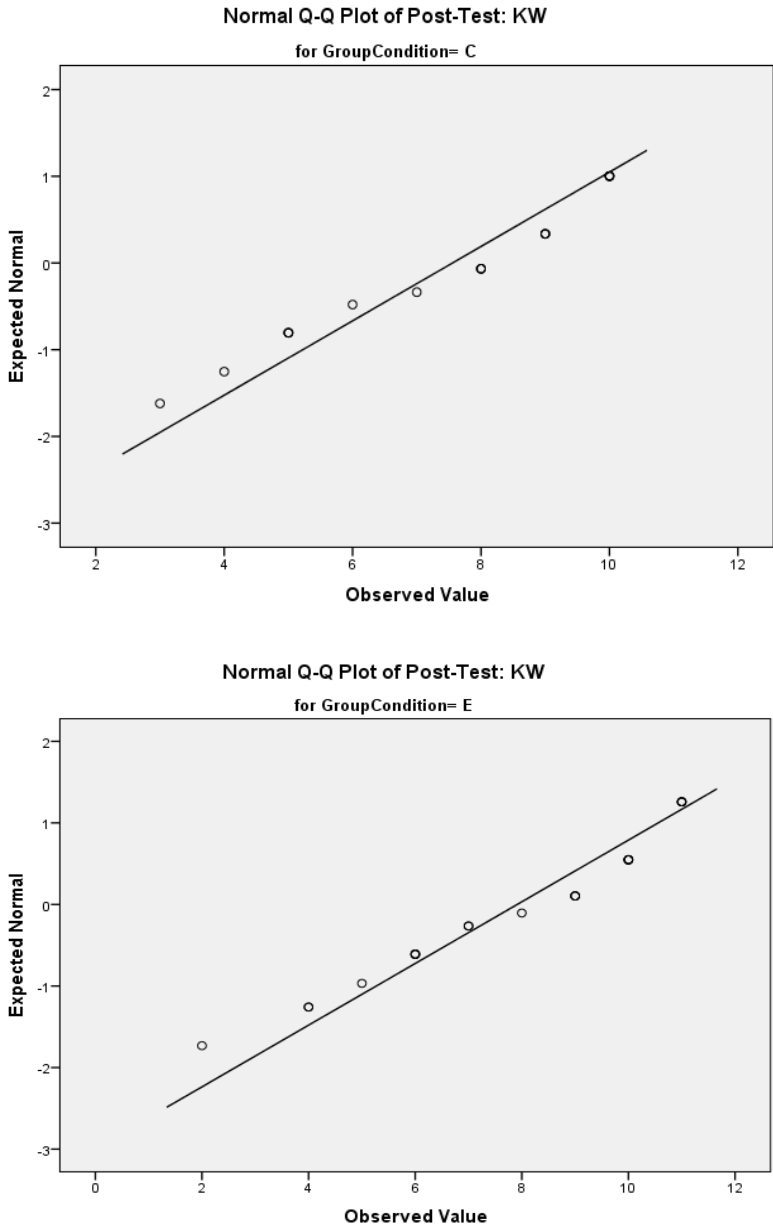
Figure 9 - *Normal Q-Q plots for the keywords score in the post-test in both groups.*

---

[18] According to Field (2014, p. 882) a Q-Q plot illustrates "quantiles of a variable against the quantiles of a particular distribution". If the values are close to the diagonal line, they share the same distribution. In the case of a normal Q-Q this distribution is normal.

By examining these Q-Q plots, it appears that the distribution of the scores in both groups approximate normal distribution to such an extent that parametric statistical tests are possible.

5.1.3 Descriptive statistics - Pre-test vocabulary size

In order to measure the pupils' receptive vocabulary the VST (Nation & Beglar, 2007) was administered as part of the post test. In group C, the mean score of the vocabulary size test was 7998 (*SD = 1250.83, SE = 294.83)* with a 95% confidence interval of 7366–8610. The minimum score was 6200 and the maximum score was 10400 (range = 4200). Figure 10 shows the scores of the vocabulary size section of the pre-test in group C. The X-axis represent the participant ID, and the Y-axis represent the vocabulary size.
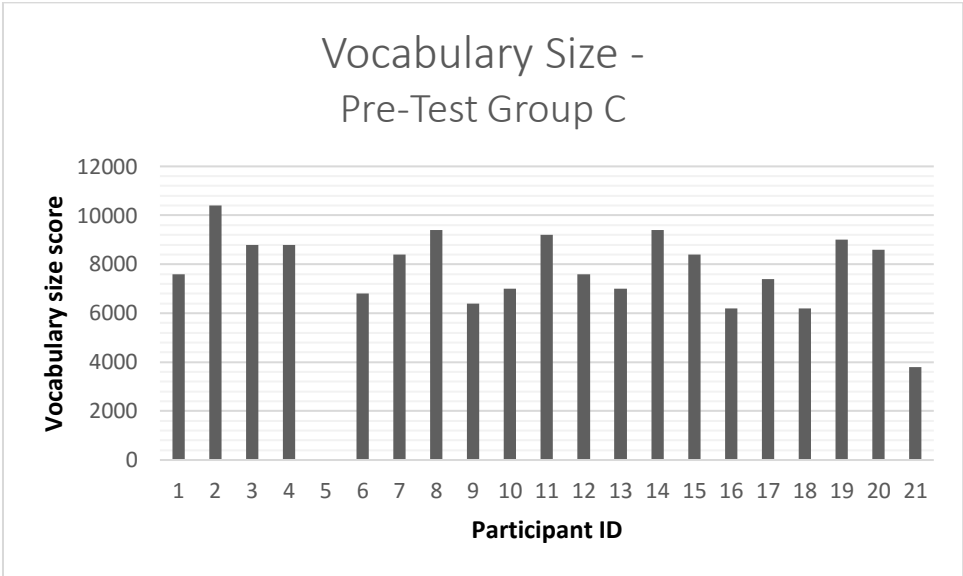
Figure 10 – *Bar chart illustrating the vocabulary size in group C.*

In group E, the mean score of the vocabulary size test was 5400 (*SD = 2351.01, SE = 490.22)* with a 95% confidence interval of 4383–6416. The minimum score was 600 and the maximum score was 9600 (range = 9000). It is important to note that 14 pupils in group E did not finish the vocabulary size test. This happened in spite of both groups receiving the same instructions and allotted time. As this test was a multiple choice test encouraging attempts at inferring the meaning of the any unknown words, with 12 questions per page, it is likely that not finishing it will adversely affect the final score. Thus, it is important to keep this in mind when comparing the results of the two groups. Figure 11 shows the scores of the vocabulary size

68

section of the pre-test in group E. The X-axis represent the participant ID, and the Y-axis represent the vocabulary size.

Figure 11 - *Bar chart illustrating the vocabulary size in group E.*

In order to measure the distribution of scores in both groups on the vocabulary test, a Shapiro-Wilk test of normality was conducted. The results of this test are summarized in Table 5:

Table 5 - Shapiro-Wilks test of normality for the vocabulary size scores in both groups.

Group	Shapiro-Wilk		
	Statistic	*df*	Sig.
C	.952	18	.452
E	.981	23	.920

In light of these results, the null hypothesis is not rejected and the data distributions are likely normal.

5.1.4 Descriptive Statistics – Concept Understanding

In order to assess pupils' understanding of the key concepts found in the global warming text, a ten-item multiple choice test was conducted during the post-test and delayed post-test. The maximum score was 10. In group C, the mean score of the post-test was 6.67 (*SD = 1.57, SE = .370)* with a 95% confidence interval of 5.89–7.45. The minimum score was 4 and the maximum score was 10 (range = 6). On the delayed post-test, the mean score increased to 7.17 (*SD = 1.51, SE = .355)* with a 95% confidence interval of 6.42–7.92. The minimum score was 4 and the maximum score was 10 (range = 6). Figure 12 shows the scores of the concept

69

understanding component of the post-test and delayed post-test in group C. The X-axis represent the participant ID, and the Y-axis represent the concept understanding score.
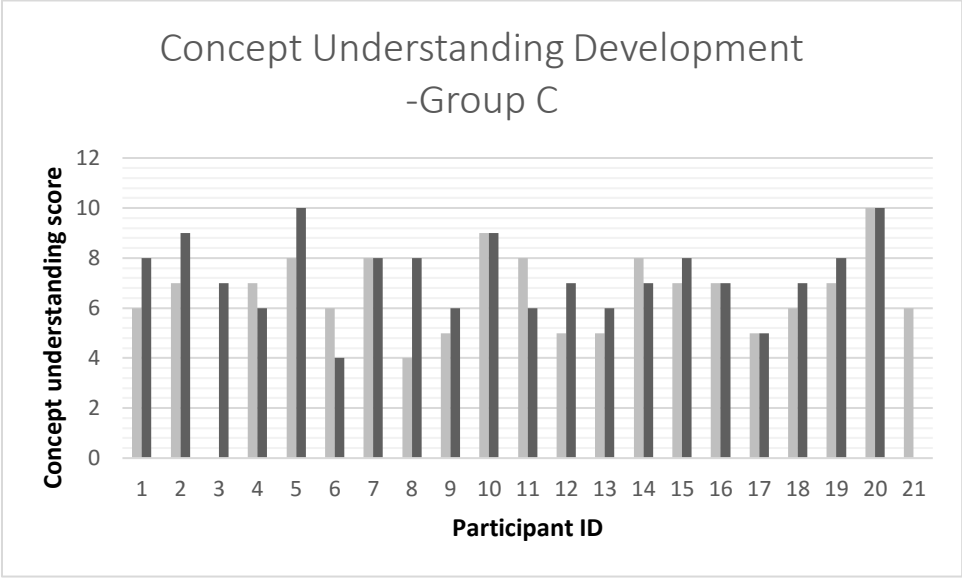
Figure 12 - *Bar chart illustrating the concept understanding score development in group C (pre-test light grey, post-test dark grey).*

In group E, the post-test mean score was 6.09 (*SD = 2.04, SE = .426)* with a 95% confidence interval of 5.20–6.97. The minimum score was 2 and the maximum score was 10 (range = 8). In the delayed post-test, the mean score was reduced to 5.70 (*SD = 2.53, SE = .528)* with a 95% confidence interval of 4.60-6.79. The minimum score was 1 and the maximum score was 10 (range = 9). Figure 13 shows the scores of the concept understanding section of the post-test in group E. The X-axis represent the participant ID, and the Y-axis represent the concept understanding score.
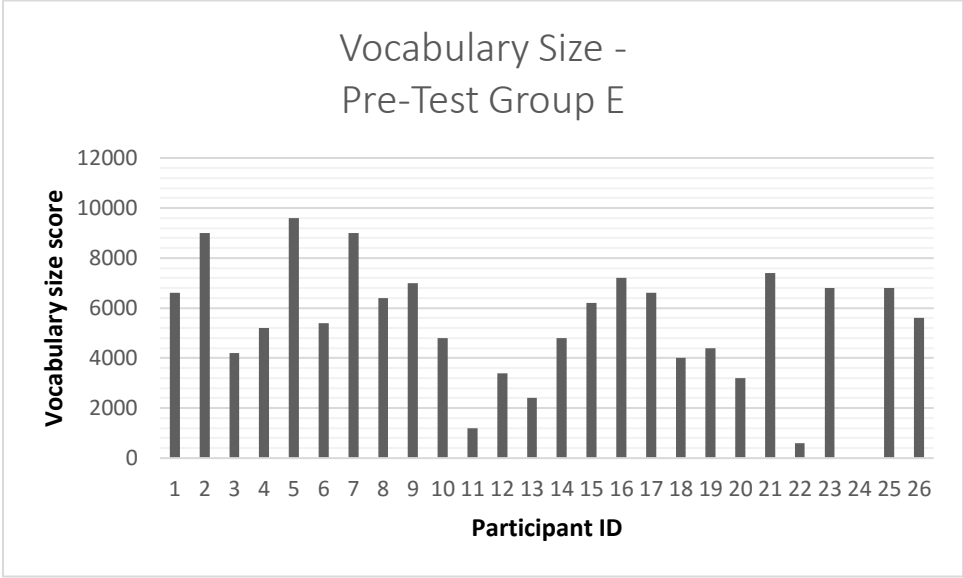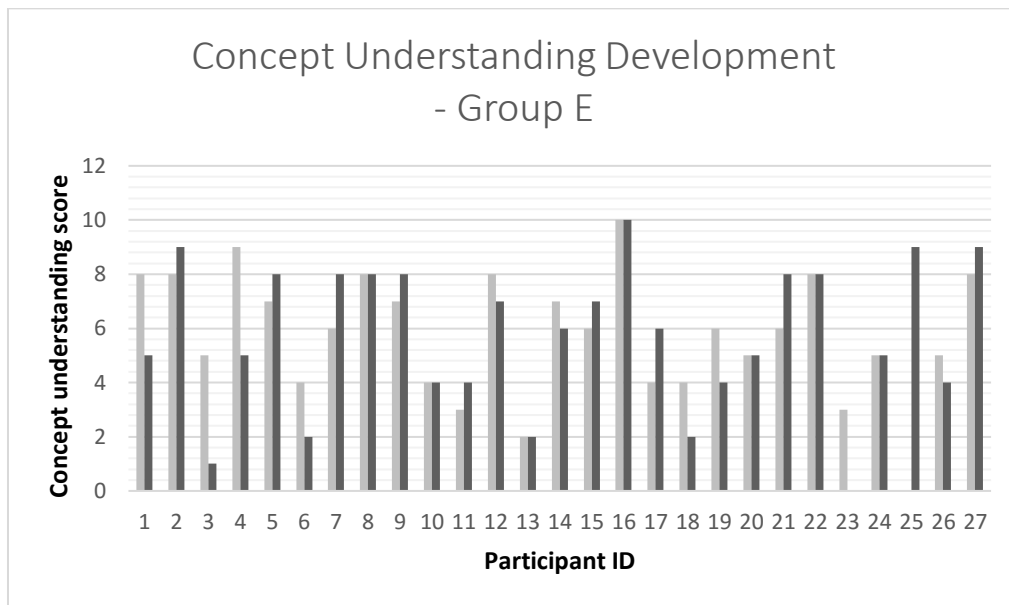
Figure 13 - *Bar chart illustrating the concept understanding score development in group E (pre-test light grey, post-test dark grey).*

In order to evaluate the distribution of the scores for all the concept understanding tests in both groups, a Shapiro-Wilk test of normality was conducted. The results of this test is summarized in Table 6.

Table 6 - Shapiro-Wilks test of normality assessing the concept understanding test scores.

Test/Group	Shapiro-Wilks		
	Statistic	*df*	Sig.
Post-test Concept Understanding C	.956	18	.522
Post-test Concept Understanding E	.969	23	.652
Delayed Post-test Concept Understanding C	.965	18	.693
Delayed Post-test Concept Understanding E	.941	23	.188

In light of these results, the null hypothesis is not rejected and normal distribution can be assumed.

5.1.5 Descriptive statistics summary

This section has presented the descriptive statistics used to describe the test scores for all the components of the pre-test, post-test, and delayed post-test. Table 7 provides a summary of all the descriptive statistics.

Table 7 – Summary of descriptive statistics.

Test	Mean	SD	95% confidence intervals
Pre-test contextual inferences group E	7.72	3.53	6-26–9.18
Pre-test contextual inferences group C	8.5	3.51	6.85–10.15
Pre-test vocabulary size group E	5400	2351.01	4383–6416
Pre-test vocabulary size group C	7998	1250.83	7366–8610
Pre-test keywords group E	4.61	2.31	3.45-5.77
Pre-test keywords group C	6.55	2.37	5.30–7.59
Post-test keywords group E	7.91	2.64	6.77–9.06
Post-test keywords group C	7.56	2.33	6.40–8.72
Post-test concept understanding group E	6.09	2.04	5.20–6.97
Post-test concept understanding group C	6.67	1.57	5.89–7.45
Delayed post-test keywords group E	8.15	2.37	7.21–9.09
Delayed post-test keywords group C	7.90	1.97	6-98–8.82
Delayed post-test concept understanding group E	5.70	2.53	4.60–6.79
Delayed post-test concept understanding group C	7.17	1.51	6.42–7.92

5.2 Inferential statistics

This section will present and discuss the findings of the inferential statistics used to answer the research questions, and to evaluate and address any pre-existing group differences on the pre-test. The following research questions will be explored in this section:

Q-1. Does using teacher-initiated explicit vocabulary teaching in the form of identification and teaching of keywords found in the subject text provide a higher level of subject understanding in the experiment group compared to the control group?

Q-2. Does using explicit vocabulary teaching in the form of identification and teaching of keywords provide a higher degree of vocabulary item acquisition in the experiment group compared to the control group?

Q-3. Is there a relationship between pupils' ability to infer word meaning from context and their vocabulary size?

Q-4. Does pupils' ability to infer word meaning from context have an effect on the effectiveness of using teacher-initiated explicit focus on vocabulary in the form of keyword identification?

Each research question will be explored using either independent means t-test, Pearson's correlation, or simple linear regression modelling.

72

5.2.1 Assessing pre-test group similarity

Because this study is based on a quasi-experiment design, there has been no random group assignment. Thus, there is a chance of there being pre-existing group differences on some critical factor. In order to address this issue, three individual differences between the pupils (contextual inferences, keywords, and vocabulary size) were assessed during the pre-test, as described in section 4.7.1. Because the distribution of the pre-test scores is very close to normal, the most robust test for checking any group differences is an independent means t-test (Larson-Hall, 2012, section 3). The independent means t-test assesses the difference between the mean scores of two groups in probabilistic terms. The formal null hypothesis is that there is no difference between the groups. A t-test results in a t-value, where "the higher the t-value, the more likely it is that the difference between groups is not a result of sampling error" (Dancey & Reidy, 2014, p. 221). Dancey & Reidy (2014, p. 221) define the p-value as "the likelihood of this arising by sampling error". If $p < 0.05$, the null hypothesis is rejected and it is likely that the difference between the means are statistically significant.

Perhaps the most important potentially pre-existing difference between the groups is the knowledge of the keywords that are being tested throughout the study. If such a difference exists, it will affect how the results of the post and delayed post-tests are interpreted. The independent means t-test found that there was a statistically significant difference between the groups (t = 2.25, p = 0.029, df = 42)[19]. The results of the independent means t-test are presented in Table 8, and a boxplot representation of the scores can be found in Figure 14. A boxplot provides a visual representation of the range of scores in a sample. According to Larson-Hall (2012, section2, final paragraph) it "indicates the median score by the horizontal black line in the box, and indicates the range of scores of the groups as well". Additionally, "the box contains the middle 50 % of the scores of the groups". Thus, it is an excellent way of representing the differences between two groups visually.

---

[19] df is an abbreviation of degrees of freedom, and in the t-test it is a result of subtracting 1 from both groups and adding the numbers together.

Table 8 - Independent means t-test pre-test keywords

T	*df*	Sig. (2-tailed)	Mean difference
2.25	42	.029	1.68
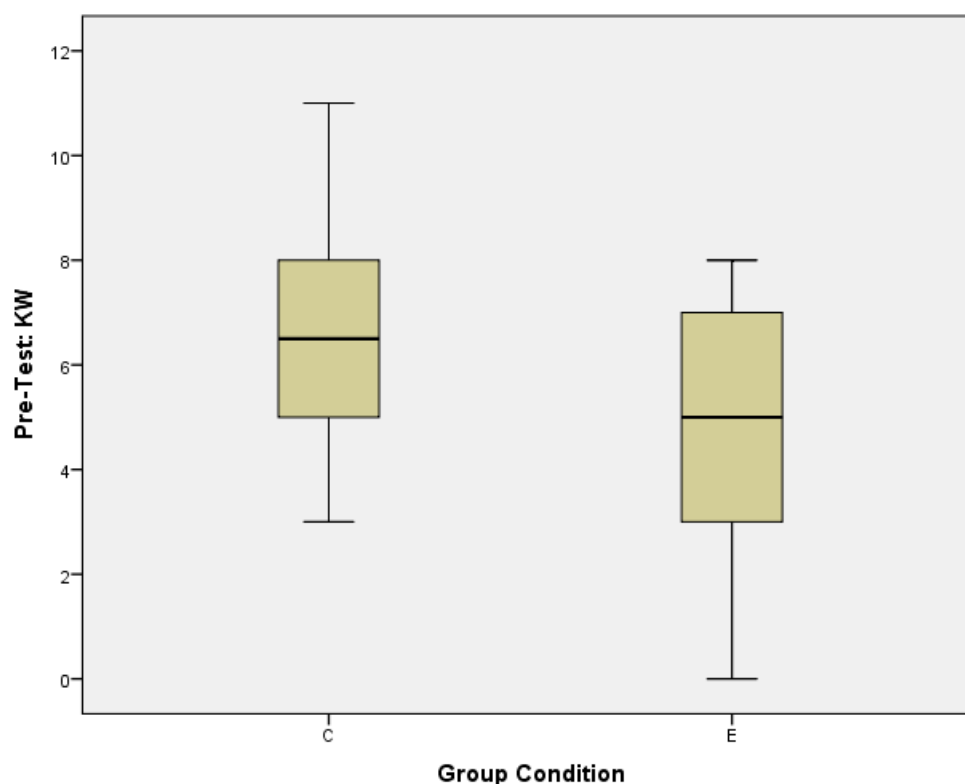
Figure 14 - *Boxplot illustration of pre-test keywords score in both groups.*

In light of these findings, it appears that group C has a higher pre-existing understanding of the keywords that are tested during this inquiry. This must be taken under consideration when the keywords test results on the post-test and delayed post-test are analyzed.

Another individual difference which was tested during the pre-test was pupils' ability to infer word meaning from context. An independent means t-test found that there was no statistical difference between the groups (t = .738, p = .465, df = 42). The finding of the independent samples t-test are found in Table 9, and a boxplot representation of the group scores are found in Figure 15.

Table 9 - Independent means t-test analysis on pre-test contextual inferences scores.

T	*df*	Sig. (2-tailed)	Mean difference
.738	42	.465	.780

74
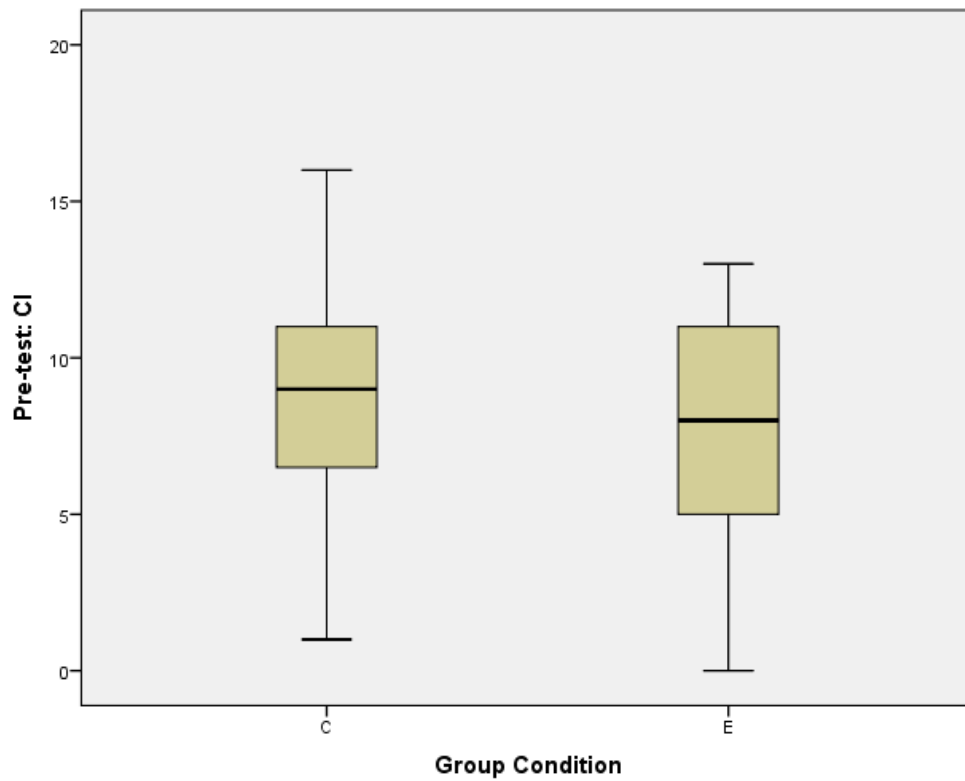
Figure 15 - *Boxplot illustrating the contextual inferences score in both groups.*

In light of this analysis, the formal null hypothesis is not rejected, and it is possible to assume that there is no difference between the groups.

Finally, an independent means t-test was conducted in order to measure any group differences on the vocabulary size test. The independent means t-test found that there was a statistical difference between the groups (t = 4.08, p < .001, df = 42). The results of the independent means t-test are presented in Table 10, and a boxplot representation of the scores can be found in Figure 16. The *21 score in group C indicates an extreme score.

Table 10 - Independent means t-test on the pre-test vocabulary size test scores.

T	*df*	Sig. (2-tailed)	Mean difference
4.08	42	< .001	2308
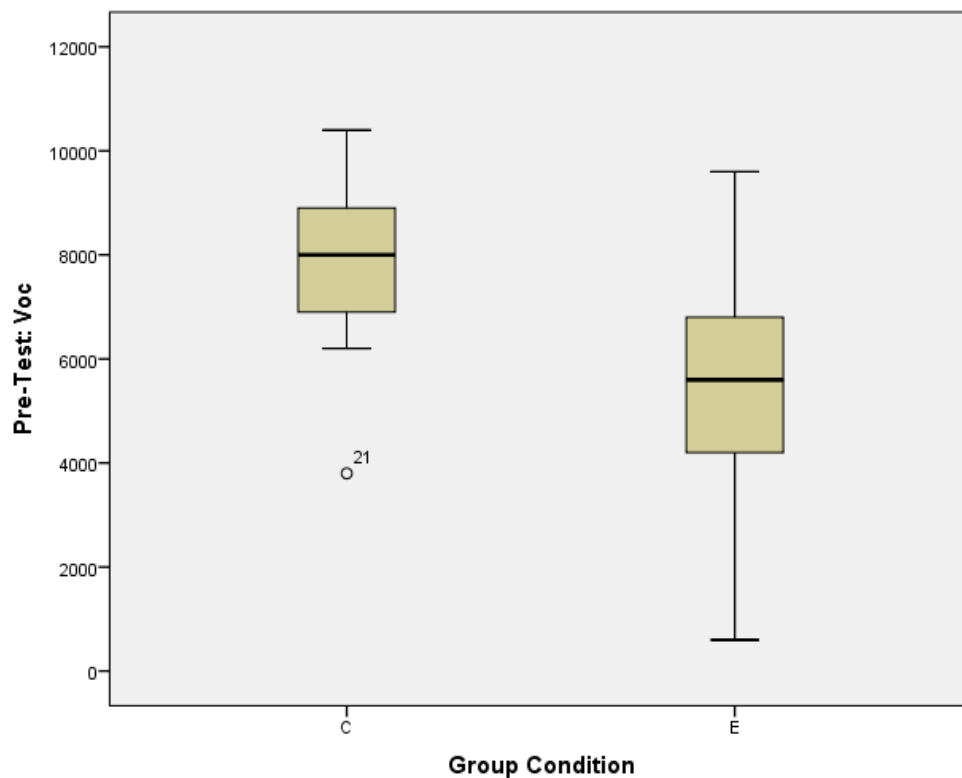
Figure 16 - *Boxplot illustrating the pre-test vocabulary size test scores in both groups.*

In light of this analysis, the formal null hypothesis is rejected and it is reasonable to conclude that there is a difference between the groups. However, as was noted in section 5.1 a large number of the pupils in group E did not finish the vocabulary size test. Had this not been the case, it is possible that the differences would have been substantially smaller. Thus, is not possible to conclude that the group differences found are not a result of the pupils in group E completing a smaller portion of the vocabulary size test. Nevertheless, this group difference must be kept in mind when analyzing the results on the post-test and delayed post-test.

5.3 Does using teacher-initiated explicit vocabulary teaching in the form of identification and teaching of keywords found in the subject text provide a higher level of subject understanding compared to the control group?

In order to explore the first research question of this study, an independent means t-test was conducted on the concept understanding components of the post-test and delayed post-test. In order to explore this question, a hypothesis was formed stating that using explicit vocabulary teaching in the form of identification and teaching of keywords will result in a

higher understanding of the text about global warming and climate change in the experiment group compared to the control group (H1).

In the post-test, the independent means t-test found that there was no difference between the groups (t = 1.324, p = .192, df = 44). This suggests that the explicit focus on vocabulary did not lead to a higher understanding of the key concepts found in the global warming text, and neither group performed better than the other. However, the independent means t-test on the delayed post-test found that there was a difference between the groups (t = 2.262, p = .029, df = 44). In the delayed post-test, the mean score in group C had increased whereas in group E it had decreased. This suggests that the increased focus on key concepts in group C has led to a higher degree of retention compared to group E. The results of the independent means t-test for both the post-test and delayed post-test are presented in Table 11. Figure 17 and 18 show a boxplot representation of the scores in both groups.

Table 11 - Independent means T-test on the concept understanding scores on the post-test and delayed post-test.

Test				
	T	df	Sig.	Mean Difference
Post-test Concept Understanding	1.324	44	.192	.700
Delayed Post-test Concept Understanding	2.262	44	.029	1.377
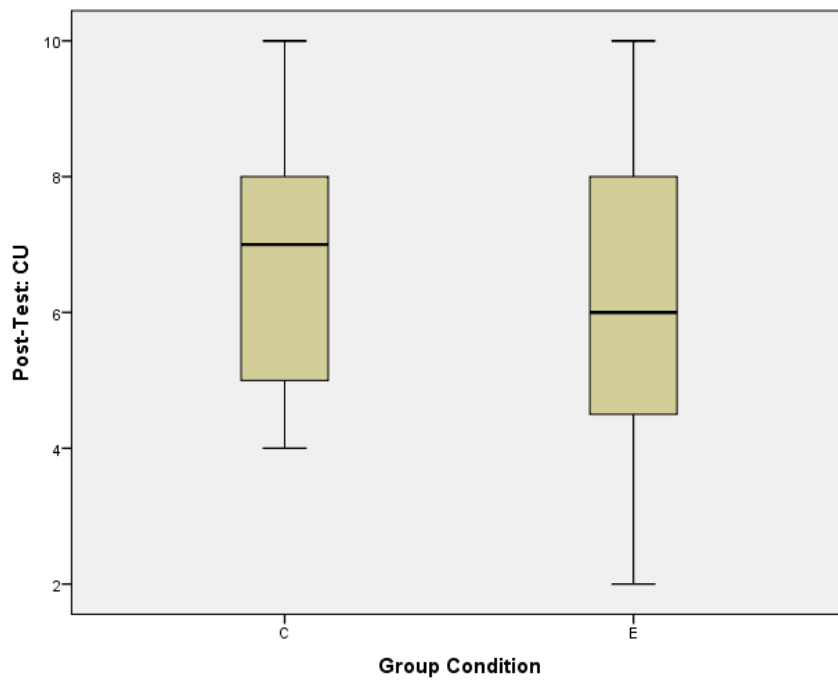
Figure 17 - *Boxplot illustrating the concept understanding scores on the post-test in both groups.*
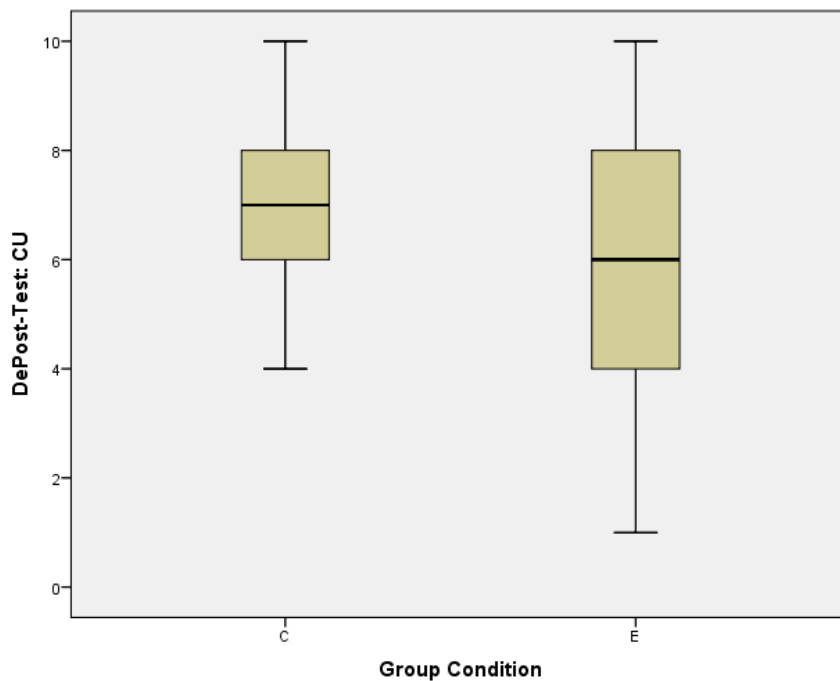
Figure 18 - *Boxplot illustrating the concept understanding scores on the delayed post-test in both groups.*

The results of the independent samples t-test are not in support of H1. Using an explicit focus on keywords in group E did not cause an increased understanding of the concepts found in the text about global warming and climate change. Moreover, on the delayed post-test group

C outperformed group E. However, as was noted in section the pre-tests suggest that there was a difference between the groups on both the vocabulary size test, and previous understanding of keywords test, with group C having a higher mean score. Pre-existing group differences such as these are always a possibility when conducting quasi-experiments using intact groups. In light of this, it is not possible to conclude that the difference between the groups on the delayed post-test was not a result of the pre-existing differences between the groups.

5.4 Does using explicit vocabulary teaching in the form of identification and teaching of keywords provide a higher degree of vocabulary item acquisition in the experiment group compared to the control group?

The second research question in this study was explored using an independent means t-test on the keywords component of the post-test and delayed post-test in order to assess group differences. In order to explore this question, a hypothesis was formed stating that using explicit vocabulary in the form of keyword identification will cause a higher degree of keyword vocabulary items acquisition from the text measured on the post-test in the experiment group compared to the control group (H2).

In order to strengthen or falsify this hypothesis, it is important to keep in mind the pre-existing difference between the groups on the keywords component of the pre-test where group C's mean score was the highest (t = 2.25, p = .029, df = 44, mean difference = 1.68). Because the E group's mean score was lower on the pre-test compared to group C, it will be highly interesting if this group difference is diminished in the post test. If this is the case, the explicit focus on vocabulary will arguably have caused a higher amount vocabulary item acquisition, in support of H2.

In the post-test, an independent means t-test found that there was no difference between the groups (t = -.588, p = .559, df = 44, mean difference = -412). Group E had the highest mean score (7.96). Similar results were found in the keywords component of the delayed post-test, where an independent samples t-test found no difference between the groups (t = -.400, p = .691, df = 44, mean difference = -.254). Once again, group E had the highest mean score (8.15). A summary of the results of the independent t-test for the post-

test and delayed post-test can be found in Table 12, and a boxplot of the scores on the keywords component of the post-test and delayed post-test can be found in Figure 19 and 20.

Table 12 - Independent means t-test analysis on the post-test keywords score in both groups.

Test				
	T	*Df*	Sig.	Mean Difference
Post-test keywords	-.588	44	.559	-412
Delayed Post-test Keywords	-.400	44	.691	-.254
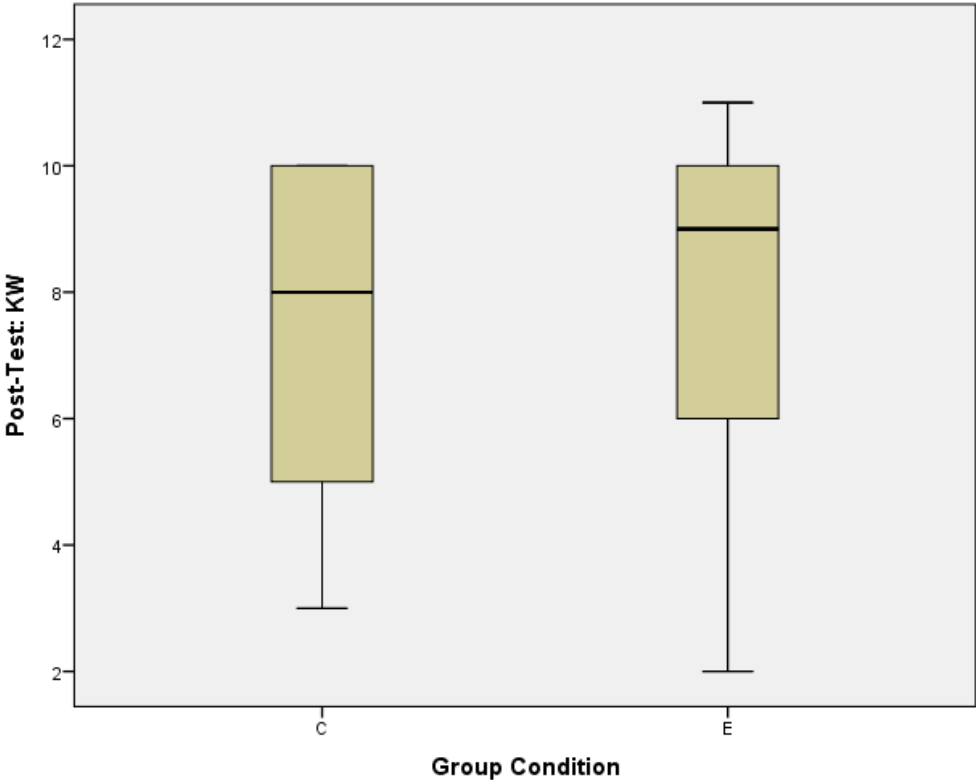
Figure 19 – *Boxplot illustrating the scores on the keywords component of the post-test in both groups.*
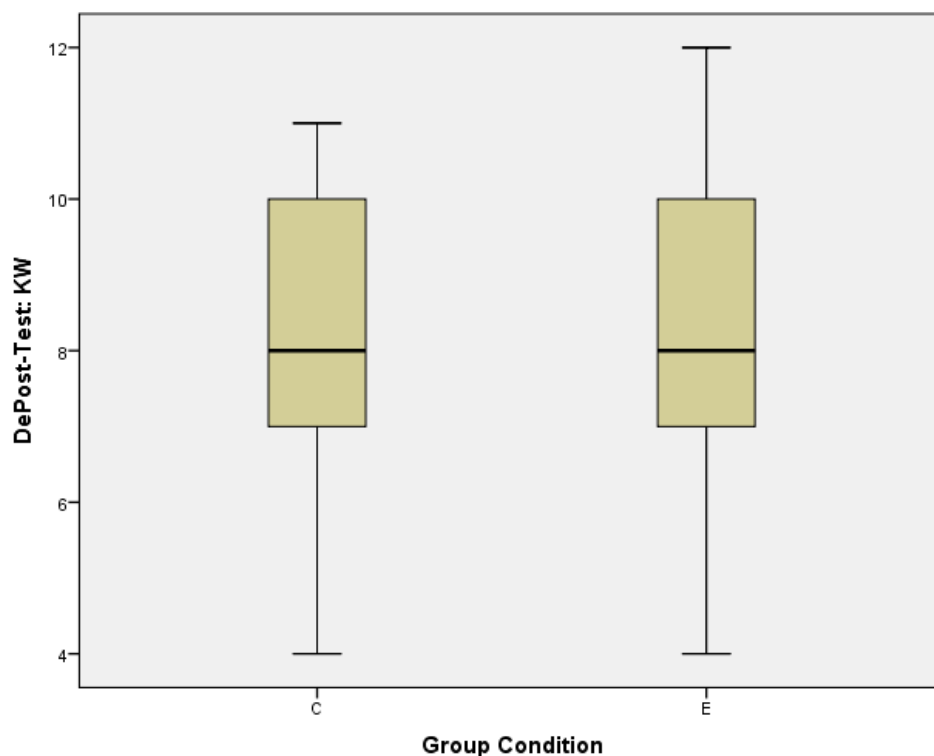
Figure 20 - *Boxplot illustrating the score on the keywords component of the delayed post-test in both groups.*

Although the independent means t-test did not find any group differences on the post-tests, the pre-existing difference between the groups discovered on the pre-test has been diminished (section 5.2.1). This suggests that the explicit focus on vocabulary in group E has caused an increased amount of keyword learning compared to group C. This is in support of the H2 hypothesis. However, it must be kept in mind that other possible explanations for this increased score growth in group E might exist.

5.5 Is there a relationship between pupils' ability to infer the meaning of unknown words from context and their vocabulary size?

The third research question in this study was explored using Pearson's correlation. In light of previous research, a hypothesis was formed postulating that there will be a positive relationship between pupils' ability to infer the meaning of unknown words from context and their vocabulary size (H3). If there is a positive correlation between the contextual inferences and vocabulary size components in the pre-test, this hypothesis is supported.

A Pearson correlation coefficient of r = .56, p <.001 (N=45) was found between the scores on the contextual inferences test and the vocabulary size test. Pearson's r indicates the strength of the relationship between to variables. This correlation coefficient ranges from 1 to -1 where 0 indicates that there is no relationship between the two variables (Dancey & Reidy, 2014). According to Larson-Hall (2012) r > .5 is large. The effect size of a Pearson's correlation is represented by R^2. In this case, R^2 = .30 suggesting that the scores on the context inferences test explained 30% of the variance in the vocabulary size test. Figure 21 is a scatterplot showing the relationship between the vocabulary and contextual inferences sections of the pre-test. The scatterplot shows that the relationship between the two variables is linear, and relatively tightly clustered around the regression line.[20] The Loess [21] line is arguably a good fit to the regression line suggesting that the relationship between the variables is linear[22].
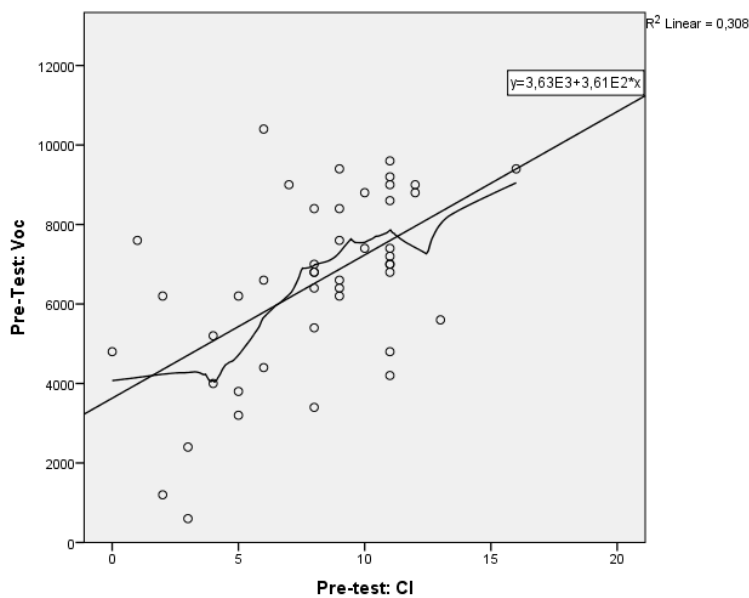
Figure 21 - *Scatterplot with regression line (straight) and Loess line (fitted) illustrating the linear relationship between ability to infer word meaning from context and vocabulary size.*

---

[20] The regression line is "a line drawn over the data which best fits all the points" (Larson-Hall, 2012, section 4, 1st paragraph.
[21] "The Loess line follows the fit of the data whereas the regression line tries to make the best fit over the entire data set" (Larson-Hall, 2012, section 4, 1st paragraph)
[22] This is one of the major assumptions of the Pearson's correlation test (Dancey & Reidy, 2014).

The results of the Pearson's correlation suggest that there is a positive relationship between contextual inferences of unknown words and vocabulary size, in support of H3. Thus, these results indicate that there is a positive relationship between pupils' ability to infer word meaning from context and their vocabulary size.

5.6 Does pupils' ability to infer word meaning from context have an effect on the effectiveness of using teacher-initiated explicit focus on vocabulary in the form of keyword identification?

In order to explore this research question, a regression model was computed for the E and C group. This model proposed that the contextual inferences score on the pre-test would function as a predictor for the keywords score on the post-test. Because the explicit focus on vocabulary in group E would help the pupils identify and understand the potentially unknown words in the global warming and climate change text, a hypothesis was made suggesting that the contextual inferences score on the pre-test would account for a high amount of the variance of the keywords score on the post-test in group C, but not in group E (H4).

Linear regression was carried out to determine the effect of pupils' ability to infer word meaning of unknown words using contextual cues on their learning of new keywords in both groups. The goal of linear regression modeling is to predict the score of one response variable (Y) by using the score of a predictor variable (X). That is, by using the line of best fit (*b*) (Dancey & Reidy, 2014, p. 380). In this model Y equals the score on the keywords section of the post-test, and X equals the score on the contextual inferences section of the pre-test. The model results in four scores: R, R^2, adjusted R^2, and standard error of the estimate. R indicates the relationship between X and Y, and is the same as a Pearson's *r* used in correlation. R^2 is an estimate of how much of the variance in Y is explained by X (Dancey & Reidy, 2014). According to Dancey & Reidy (2014, p. 394) "R^2 is too optimistic, as the line of best fit is based on a sample, not the population". Additionally, as the goal is to generalize to the whole of the population, "[a]djusted R Square adjusts the figure to give a more realistic estimate". The standard error of the estimate "gives us a measure of how accurate our estimation is likely to be" (Dancey & Reidy, 2014, p. 395). The regression model for both groups can be found in Table 13.

83

Table 13 – Summary of the linear regression models for group C and E, and ANOVA table.

Group Condition	Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics R Square Change
C	1	.591[a]	.349	.311	1.884	.349
E	1	.322[a]	.103	.063	2.510	.103

Group Condition	Model		Sum of Squares	df	Mean Square	F	Sig.
C	1	Regression	32.409	1	32.409	9.133	.006
		Residual	60.328	17	3.549		
		Total	92.737	18			
E	1	Regression	15.983	1	15.983	2.536	.166
		Residual	138.642	22	6.302		
		Total	154.625	23			

In group C, adjusted R^2 = .311 (p = .008, df = 17). This indicates that the contextual inferences score accounted for 31% of the variance in the keywords score on the post-test. In group E, adjusted R^2 = .063 (p = .126, df = 22) suggesting that the contextual inferences score accounted for 6.3% of the variance in the keywords score on the post-test. In order to assess if the model is statistically significant, it is necessary to check the p-value found under Sig. in the ANOVA table. This provides an indication of whether the prediction is better than what could be expected by chance (Dancey & Reidy, 2014). In group C, it was found that the results were statistically significant (F = 9.13, p = .008, df = 17), suggesting that the model predicts the score of Y better than would be expected by chance. In Group E, the results were not found to be statistically significant (F = 2. 54, p = .126, df = 22) suggesting that the predictions of the model would not be better than chance. The slope of the line of best fit (b) and intercept (a) can be found in Table 14.

Table 14 – Coefficients in both linear regression models (group E and C).

Group Condition	Model		Unstandardized Coefficients		Standardized Coefficients	
			B	Std. Error	Beta	T
C	1	(Constant)	4.350	1.137		3.827
		Pre-test: CI	.382 (*b*)	.126	.591	3.022
E	1	(Constant)	6.093	1,231		4.950
		Pre-test: CI	.231 (*b*)	.145	.322	1.593

Group Condition	Model		Sig.
C	1	(Constant)	.001
		Pre-test: CI	.008
E	1	(Constant)	.000
		Pre-test: CI	.126

This table suggests that in group C the keywords score would increase by .382 for every one point increase in the contextual inferences score. In group E, a one point increase in the contextual inferences score would result in a .231 increase in the keywords score. Table 15 provides an overview of the residuals for the regression models in the control and experiment group. According to Dancey & Reidy (2014, p. 385) "[t]he differences between the actual scores and predicted scores are called *residual*s". In other words, the residual statistics provide an indication of how well the line of best fit predicts the score of the response variable (Y).

Table 15 - Residuals Statistics for both models (group E and C)

Group Condition		Minimum	Maximum	Mean	Std. Deviation	N
C	Predicted Value	4.73	10.46	7.53	1.342	19
	Residual	-3.552	2.594	.000	1.831	19
	Std. Predicted Value	-2.083	2.188	.000	1.000	19
	Std. Residual	-1.885	1.377	.000	.972	19
E	Predicted Value	6.09	9.10	7.88	.834	24
	Residual	-5.099	4.214	.000	2.455	24
	Std. Predicted Value	-2.138	1.468	.000	1.000	24
	Std. Residual	-2.031	1.679	.000	.978	24

In the experiment group, the minimum and maximum residuals are between -3.552 and 2.594. This is smaller than in the control group where the minimum value was -5.099 and the maximum value was 4.214. This suggests that the line of best fit is better at predicting the scores on the response variable in the group model computed for group E compared to the model computed for group C.

In light of the regression models computed for the experiment and control groups, the findings indicate that the context inferences score was a good predictor in the control group (C), but not in the experiment group (E). This is in support of H4, and suggests that the explicit focus on vocabulary in the experiment group has mitigated the effect of one individual difference. However, because the independent means t-test conducted on the vocabulary size test component of the pre-test indicated that there was a difference between the two groups, there is a chance that that the lower mean score in the experiment group has functioned as a confounding variable. In order to explore this, a separate model for both groups using both the contextual inferences and vocabulary size test scores on the pre-test as predictors for the keywords post-test score was made. However, the addition of the vocabulary size test score did not improve the regression models in either group condition. In the experiment group, adjusted R^2 = .257 (F = .158, p = .696, df = 16) and in the experiment group adjusted R^2 = .035 (F = .373, p = .548, df = 21). This suggests that the vocabulary size did not play an important part in the prediction of the keywords score in the post test.

CHAPTER 6 – DISCUSSION

This chapter will discuss the findings of this study. The main focus will be on discussing the practical and theoretical implications of the results. This will be done in order to achieve the two aims of this master's thesis. The first aim was to explore the usefulness of the teacher-initiated identification and explicit focus on keywords as an approach to provide the pupil's with an adequate understanding of an academic topic, and acquisition of new vocabulary items. The second aim was to explore the usefulness of this approach in relation to pupils' ability to infer word meaning from context and vocabulary size. This discussion will be divided into four sections exploring the theoretical and practical implications of the results, as well as proposals for new areas of research. The first section will identify and address the weaknesses and limitations of the methods that have been used. These weaknesses and limitations must be kept in mind when the theoretical and practical implications are being discussed.

6.1 Weaknesses and limitations

Because this study has been conducted in a naturalistic setting using intact groups, it has been identified as an interventionist quasi-experiment (Loewen & Philp, 2012). Even though this form of instructed language acquisition research has been identified by Loewen & Philp (2012) as having several strengths compared to other forms of research, the most important weaknesses and limitations of this master's thesis are likely embedded in the research design itself. Some of these, e.g. group differences, classroom noise level, deviation from teaching schedule, and deviation from teaching plans, were identified and addressed in section 4.7. These were all proposed as potentially confounding variables, and were addressed during the design of the study. However, although several steps were taken in order to achieve a high degree of validity and reliability, there are still several weaknesses and limitations that must be taken into consideration.

As this investigation was conducted in a classroom setting, the amount of experiment control has been limited, in spite of the validity and reliability measures that were part of the research design. Because of this, it is impossible to rule out all potentially confounding variables. The only possible way to account for these confounding variables would have been to conduct the quasi-experiment in a non-naturalistic setting. However, this would have added a high degree of artificiality to the teaching environment. Moreover, it would have made it very difficult to use pupils from the Norwegian lower secondary school system. Thus, this is a

weakness that has not been possible to fully account for during this research project. In light of this, it is important to keep in mind that the results of this investigations can have been affected by confounding variables that have not been discovered or accounted for. Nevertheless, the use of a naturalistic setting has also added a high degree of authenticity to the teaching activities and methods that have been explored. Additionally, it has allowed for the use of age-appropriate pupils. This might be thought of as beneficial for the exploration of practical implications that is one of that main goals of this thesis. Thus, it should make this research more useful for practicing teachers.

A second weakness of this research design is the use of intact groups. This is a result of the naturalistic setting, and is a major weakness for most quasi-experimental designs. Because the use of intact groups could lead to pre-existing group differences, several steps were taken in order to assess group similarity. As was identified in section 5.2, the two groups had pre-existing differences related to their vocabulary size, and pre-existing knowledge of keywords. Although these differences have been discovered, it is not possible to rule out that other group differences also exist. It is impossible to check for and rule out all potential differences between the groups; in the case of the t-test analyses the only possible way to address this weakness completely would be to use random group assignment. However, doing this in a naturalistic setting would make it very difficult to find a teacher and pupils who are willing to break up the existing classroom environment for the duration of the quasi-experiment. This would also add a high degree of artificiality to the study, and function as a confounding variable. Additionally, it could lead to a higher amount of classroom noise. It would also be possible to model the individuals as a random effect in a mixed effects regression model. Nevertheless, this is a major weakness of this study, and should be taken into consideration during the discussion of theoretical and practical implications.

A third weakness that has partially been a result of the naturalistic setting and quasi-experiment design is the sample size. The total number of participants that have been participating in all of the three tests is 41, with 43 participating in the pre-test and post-test (19 in group C and 24 in group E). This is a moderately small sample size and the 95 % confidence intervals are therefore high (se section 5.1.5). Thus, it is difficult to generalize the findings to the overall population. This is partly a result of the typical classroom size in the Norwegian school system, with classes rarely exceeding 30 pupils. Additionally, pupils fall ill,

88

or have other appointments which might cause them to be absent. Although it would be possible to use time from other subjects in order to take the tests at different times, this would arguably have caused the pupils to lag behind in other subject. Thus, this step was not taken. It should be noted that small sample sizes in instructed language acquisition is not unheard of (e.g. Zaid, 2009).

One possible way to address this weakness would have been to use more than one class for each group condition, effectively doubling the sample size. However, this would require the use of a different school and teacher, and most likely cause the workload to exceed that which is possible for one master student. Nevertheless, this step is possible, and is further addressed in section 6.4. Because of this, the findings are limited to this exact setting and these groups. Arguably, this is the major limitation of this thesis. Thus, although the theoretical and practical implications in section 6.2 and 6.3 are being discussed in general terms, it is imperative that this limitation is being kept in mind. Nevertheless, I am confident that this inquiry can be used as a starting point in order to explore the usefulness of teacher-initiated explicit focus on vocabulary, in relation to pupils' ability to infer the meaning of unknown words from context. In light of this, this thesis can be thought of as a pilot-study.

A weakness that is unrelated to the quasi-experimental design is the lack of pilot testing of the instruments that have been used during this inquiry. As was identified in section 4.4, there is a limited amount of time available for conducting a master's thesis. Because of this, spending time on pilot testing would be unfeasible. Thus, I decided to use pre-made tests where possible. This would also allow for a greater emphasis on the theoretical implications of the results, and make it possible to relate them to previous research using the same testing instruments. However, this was not possible for the keywords and concept understanding component of the testing instruments. These had to be made based on the National Geographic text that was used. The details of how these test were constructed can be found in sections 4.4.3 and 4.5. Because these testing instruments were not piloted, there is a chance that some unknown weakness has affected the results. Nevertheless, this would affect both group conditions similarly and should therefore not have a very great impact on the inferential statistics used to answer Q-1 and Q-2. Even so, it should be noted that these tests were not piloted during the discussion of theoretical and practical implications sections of this thesis.

Finally, I want to draw some attention to the contextual inferences test used in this study. There is a chance that the pupils have also drawn upon morphological cues when they have tried to infer the word meaning. I will argue that the local contextual cues will play a much more major role in this text compared to morphological cues, as the meaning of most of the unknown words are not easily recognized by using morphological cues (e.g. *stormed*, *apex*, *blayned*, etc.). Nevertheless, it is possible that the contextual inferences test used in this inquiry has also been influenced by pupils' ability to use morphological cues.

6. 2 Theoretical implications

This section will discuss the theoretical implications of the results in relation to the two aims of this thesis. The focus will be on how these findings are related to previous research, and how they can possibly expand and elaborate on the body of common knowledge. However, before the theoretical implications can be discussed, it is necessary to examine how the main findings of this study relate to the theoretical framework presented in Chapter 3. Because a vast body of research has been compiled examining vocabulary size, and the vocabulary needed for various uses of the L2, a natural starting point is to examine the result of the vocabulary size test in relation to this previous research.

6.2.1 Vocabulary size compared to previous research

As was pointed out in section 5.1, the mean scores for the two groups were 7998 in group C and 5400 in group E. In group E, 14 pupils had not finished the vocabulary size test (*VST)*, and two scores should be considered outliers (600 and 1200). Thus, there is a chance that the mean score in this group is lower than it should have been. Regardless, because the VST was developed, and has been used extensively, by Nation, it is possible to relate the results of this study to those conducted by him. As was asserted by N. Schmitt & D. Schmitt (2014), Nation (2006), and Schmitt et al. (2011), L2 readers would need to understand 98% of the vocabulary in a text, and a vocabulary size of around 8000-9000 word families, in order to read authentic texts. In group C, 10 pupils had a VST score indicating that their vocabulary size was between 8000–11000 word families. This represented almost half of the total group. In group E, only 3 pupils had scores indicating a vocabulary of more than 8000 word families. However, as was noted in section 3.3, it is not possible to expect secondary school pupils in the Norwegian school system to have a native-like vocabulary size. And the VST scores suggests that their vocabulary should be large enough to read and understand a range of different English texts.

This has implications for the suitability of the text about global warming and climate change that was used during this inquiry. In light of the VST mean scores for both groups, I argue that the text about global warming and climate change is well suited for the pupils in this inquiry.

It is important to note that the VST only provides a measurement of pupils' receptive vocabulary size. As was identified in section 3.4, several levels of word knowledge are being considered by researchers (e.g. Laufer & Goldstein, 2004; Macaro, 2003; Nation et al. 2015; Schmitt, 2010; Schmitt & Schmitt, 2012). Because of this, it is not possible to make any suggestions of the pupils' productive vocabulary size, or vocabulary depth. Furthermore, the VST uses a multiple choice format with four alternatives. This means that there is a good chance that the pupils' vocabulary score is exaggerated, because one should expect the pupils to get one fourth of the answers correctly simply by guesswork. It should be noted that Nation & Beglar (2007) encourage guessing and inferences in the VST, postulating that this might draw on some form of implicit or subconscious vocabulary knowledge. Nevertheless, the purpose of this thesis is not to explore the vocabulary size of Norwegian secondary school pupils. Moreover, in light of the theoretical framework presented in section 3.3, one can argue that receptive vocabulary plays a more important role in reading comprehension compared to productive vocabulary.

6.2.2 The effect of using explicit focus on vocabulary

The first aim of this thesis was to investigate the effect of using teacher-initiated explicit focus on vocabulary on both pupils' understanding of the global warming text, and on their acquisition of new vocabulary items. The t-test analysis in section 5.4 found no group difference on the pupils' understanding of key concepts component of the post-test. This indicates that the explicit focus on vocabulary in group E did not help the pupils with understanding the global warming text more effectively than the increased focus on content in group C. Moreover, the t-test analysis on the delayed post-test indicated that there was a difference between the groups, with group C having the highest mean score. This means that

the H1 hypothesis was not supported[23]. Thus, the results seem to indicate that the increased focus on content in group C was more effective than the focus on vocabulary in group E in terms of *retention*. Nevertheless, the explicit focus on vocabulary did not lead to a decreased amount of concept understanding on the post-test. Thus, the results of this investigation are in support of Bauman et al.'s (2003) findings. However, it should be noted that Bauman et al.'s (2003) study was not conducted in an L2 setting.

I propose three possible explanations for these results. (1) Perhaps the increased focus on vocabulary in group E caused a higher focus on the keywords when the pupils were working on the text about global warming and climate change. This focus on the keywords may possibly cause the pupils to pay less attention to the key concepts in the text. Alternatively, (2) the increased focus on key concepts during the teacher-led PowerPoints may have been more effective than the keywords oriented presentation in group E. Finally, (3) it is possible that a combination of the two methods used in the different group conditions may be more effective than using only one.

Additionally, several of the weaknesses identified in section 6.1 can be thought of as highly influential on the results. In section 5.2, it was discovered that the two groups differed on the pre-test scores measuring vocabulary and previous knowledge of the keywords. The mean score for group C was the highest on both test components. It is therefore possible to expect that their understanding of the key concepts component on the post-test should be higher than group E. However, this was not the case. In light of this, one can make a strong case for the explicit focus on vocabulary having negated these pre-existing group differences, resulting in no differences between the groups on the post-test. Still, in light of the point mentioned above, it is not possible to draw any hard conclusions based on the data collected during this quasi-experiment. Therefore, more research into this aspect of using explicit focus on vocabulary is needed, preferably using matched groups, or random group assignment. Nevertheless, the teacher-initiated focus on vocabulary did lead to a key concept

---

[23] Using explicit vocabulary teaching in the form identification and teaching of keywords will result in a higher understanding of the text about global warming and climate change in the experiment group compared to the control group.

understanding in group E which was similar to group C in spite of the pre-existing group differences.

The results of the independent t-test analyses on the keywords development of the pre-test, post-test, and delayed post-test are more favorable towards the use of teacher-initiated focus on vocabulary. The pre-test scores on the keywords component indicated that there was a difference between the E and C group, with group C having the highest mean score. However, on the post-test and delayed post-test no group differences were found. Moreover, the mean score of group E on both tests was now the highest, being almost two times higher than on the pre-test (see section 5.1.5). These results suggest that the experiment group condition caused a higher degree of vocabulary item learning compared to group C. The H2 hypothesis suggested that using explicit vocabulary teaching in the form of keyword identification would cause a higher degree of keyword vocabulary items acquisition in the experiment group compared to the control group. Thus, these results are arguably in support of the H2.

These results seem to contradict some of the findings of Chun et al (2012).Their results suggested that leaving pupils to learn words from context through extensive reading without any form of explicit focus on vocabulary caused a higher degree of retention compared to a group using *paired associate learning* (PAL). Of course, it is important to note that Chun et al (2012) used PAL and not a focus on keywords. The focus on keywords used in this thesis used a substantially larger range of vocabulary learning activities compared to PAL. Additionally, the vocabulary items were also seen in context in both groups. The findings of this inquiry instead appears to support the findings of Schmitt & Sonbul (2010), Bauman et al. (2003), and to some extent Bolger et al (2008); the focus on vocabulary did cause a higher degree of vocabulary item acquisition. These implications are also in support of Macaro's (2003) assertions regarding the importance of explicit focus on vocabulary.

In light of the weaknesses identified in section 6.1, these results must be treated with a high degree of caution. It is perfectly possible that the increased amount of word learning in group E was caused by some currently unknown variable, and not the explicit focus on vocabulary. Moreover, a competing explanation can be that the lower basepoint in group E meant that the threshold for learning new words was lower in group E. This is supported by the contextual inferences test which did not differ between the two group conditions.

Nevertheless, the results of this inquiry suggest that the explicit focus on vocabulary was more effective, and caused a higher degree of vocabulary item acquisition than focusing on the content of the global warming and climate change text.

Although the results of this investigation suggest that the explicit focus on vocabulary did not lead to an increased understanding of the key concepts in the global warming and climate change text, it did cause a higher degree of word learning. This suggests that focus on keywords is effective in providing the pupils with the tools needed to learn new words in an authentic text. Perhaps more importantly, the increased learning of vocabulary items did not interfere with the pupils' understanding of the concepts in the text, at least not on the post-test. Thus, if the goal of the teaching sessions is to maximize vocabulary development, spending time on explicit focus on vocabulary appears to be effective.

6.2.3 The effect of individual differences – Vocabulary and contextual inferences

In order to explore the third research question, a Pearson's correlation between the contextual inferences and VST score on the pre-test was conducted. This analysis found a strong correlation between these two individual differences, in support of H3.[24] Even so, as with any correlation it is not possible to determine cause and effect. In other words, one cannot determine that being good at inferring word meaning from context leads to a higher vocabulary size. Furthermore, it is not possible to rule out the possibility of several other factors playing a crucial role in this relationship. Moreover, one possibility is that pupils with a large vocabulary size already knew the meaning of some of the words in the contextual inferences text. This is, of course, only relevant for the sentences with proper words, and does not relate to the constructed non-words (*blayned, sconziatic, peltively).* However, I will argue that the words used for this test are rare enough for this to be unlikely.

Nevertheless, a potential explanation for the correlation between these two variables is that there exists a reciprocal relationship between pupils' ability to infer word meaning from context and vocabulary size; on the one hand a large vocabulary will likely support contextual

---

[24] There will be a positive relationship between pupils' ability to infer the meaning of unknown words from context and their vocabulary size (H3)

inferences, and on the other hand being good at contextual inferences can help with vocabulary development during, for example, extensive reading. This relationship has been proposed by both de Leeuw et al. (2014) and Bolger et al. (2008), and my findings seem to support this postulation, at least in relation to local inferences. Moreover, the results are also in support of Nassaji's (2006) finding. However, it should be noted that his study was oriented towards the effect of pupils' vocabulary depth on the success of contextual inferences of unknown words.

The second aim of this thesis was to examine the effectiveness of explicit focus on vocabulary in relation to pupils' ability to infer word meaning from context. Because the focus on keywords in group E might help pupils who are struggling with making contextual inferences, the H4 postulated that the contextual inferences score on the pre-test would function as a predictor for the keywords score on the post-test in group C, but not in group E. Instead, the focus on keywords would help the pupils with understanding the potentially unknown vocabulary in the global warming and climate change text. Thus, the effect of pupils' ability to understand unknown words from context would be negated by the focus on keywords in group E.

The findings of the separate linear regression models made for the E and C group seem to support this hypothesis. As was noted in section 5.6 the pupils' score on the contextual inferences component accounted for 31% of the variance in group C compared to 6% in group E. The only difference in the treatment between the groups was the focus on keywords in group E versus increased focus on content in group C. Because there was a difference in vocabulary size between the groups, a separate model where vocabulary size was accounted for was computed with similar results. Thus, the most likely explanation for this difference is that the focus on keywords in group E has removed the effect of this individual difference.

Once again caution should be exerted in making any hard conclusions based on the findings of this study. The sample sizes in both groups are relatively small, and the method

95

used to explore H4 is relatively unorthodox[25]. One of the assumptions for linear regression is that the predictor and response variable should not be affected by a third variable. In this case, it is possible to argue that the group condition violates this assumption.[26] Nevertheless, this should only affect the extent to which these results can be generalized to the population (Field, 2014, p. 309). As was identified in section 6.1, one of the major weaknesses for this thesis is that the results are restricted to this exact setting. Thus, I will argue that the results of the two regression models are valid in this setting, and that they are in support of H4. Finally, as was noted in section 6.1 there is a possibility that some unknown crucial factor is affecting the results. Because of this, the differences between the linear regression models in the two groups can be the result of some unknown difference which has not been tested for.

In the introduction, I identified that there has been a noticeable lack of research exploring the usefulness of different methodological approaches in relation to individual differences, and those studies identified by Skehan (1989) and Robinson (2002) are not related to the use of explicit focus on vocabulary. Similarly, the contemporary research presented by Lightbown & Spada (2013) was not oriented towards pupils' ability to infer word meaning from context as an individual difference. It is therefore difficult to relate and support the findings of this study to previous ones. The results of this investigation indicates that the use of explicit focus on vocabulary can help support pupils who are not good at making contextual inferences of unknown words. However, as this is only one study with a limited sample size it is difficult to generalize these findings to the whole population. The practical implications of these results will be discussed in section 6.3.

6.3 Practical implications

In the introduction, I stated that the main justification for doing this study was a practical one. I wanted to explore the usefulness of using a teacher-initiated focus on vocabulary in relation to pupils' ability to infer word meaning from context, because an important feature of the

---

[25] H4: The contextual inferences score on the pre-test will account for a high amount of the variance of the keywords score on the post-test in the control group, but not in the experiment group.

[26] The use of linear regression in an experimental study is also uncommon. An interesting comparison between the ANOVA and linear regression can be found in Field (2014, p. 431).

Norwegian school system is the principle of adapted education. If a methodological approach using a specific set of teaching activities can help mitigate the effect of an individual difference, it can be of use to teachers who want to support pupils whose abilities in this individual difference are limited. This would be in line with adapted education, and becomes especially important when the teaching activities are based on a text which is not only difficult, but also deals with a topic that is uncommon in the English classroom, and which uses a vocabulary with a large amount of technical or academic words. This section will discuss the practical implications of the results of this inquiry.

6.3.1 The viability of teaching global warming in the English classroom

A practical aspect of this thesis that has been largely ignored up to this point is the usefulness of using a text about global warming in an English classroom. This aspect is most likely highly important for practicing teachers who are contemplating trying to use this approach. The time available to English teachers in the Norwegian school system is limited. Thus, although I demonstrated in Chapter 2 that the teaching activities used in both group conditions are supported by the learning goals found in LK06 (UDIR, 2013), spending valuable classroom time on activities that have not caused some form of learning is not an option. Because of this, I will use the first part of this section to discuss the amount of learning that has occurred during the two teaching sessions used for both groups.

The most obvious indication of learning in both groups is arguably found in the keywords tests. As was demonstrated in section 5.1, both group conditions had a marked increase in the mean scores between the pre and post-tests. This would suggest that the two teaching sessions based on the text about global warming (National Geographic, 2015) have led to a noticeable increase in the pupils' understanding of the keywords in both group conditions. Similarly, the mean score for both group conditions on the concept understanding test is 6 out of 12 points (C = 6.67; E = 6.09). Arguably, this suggests that some form of learning has occurred. However, it is up to the individual teacher to determine if these scores indicate an adequate level of text understanding.

Nevertheless, there is little doubt that the teaching activities have caused an increase in pupils' understanding of the keywords, and that they have understood some of the concepts in the global warming text. In light of this, I will argue for this form of teaching as a

venture away from the regular textbooks often used in the English classroom. Moreover, with careful planning, the teaching activities used in the English sessions can be complemented with other activities during natural science or social science sessions.

6.3.2 The use of teacher-led explicit focus on vocabulary

A major practical implication of the results of this study is related to the use of teacher-initiated explicit focus on vocabulary in the English classroom. The findings suggest that the identification and presentation of keywords had a positive effect on the growth of pupils' understanding of the keywords in group E compared to group C. Even though the same results were not found for the pupils' understanding of the key concepts, the increased vocabulary growth is highly in support of this explicit focus on vocabulary. As was noted in section 6.2 this notion is also supported by other studies (e.g. Bauman et al, 2003; Beglar et al, 2008; Schmitt & Sonbul, 2010). Thus, there is a clear indication that an explicit focus on vocabulary should be part of the English classroom. Of course, it is also important to note that other studies have suggested that some forms of explicit focus on vocabulary is not more effective than ,for example, extensive reading (e.g. Chun et al. 2012). Moreover, it is not possible to say to how often or how much time should be spent on this focus in order to maximize its effect in light of these results. Additionally, the possible synergy between focusing on vocabulary *and* content has not been explored in this study. Perhaps a combination of the two approaches used in the two group conditions during this quasi-experiment is more effective than focusing on simply one. This notion is further addressed in section 6.4.

Arguably, the most important aspect of the explicit focus on vocabulary in this study is how it appears to have negated the effect of pupils' ability to infer word meaning from context. If these findings can be generalized to the population, it would suggest that this approach can help teachers conform to the concept of adapted education. If this explicit focus on vocabulary can help pupils who are struggling to infer word meaning from context understand new words when they are engaging with difficult texts, it will be important that teachers spend time identifying, presenting, and working with these keywords in order to support their language development. A different but also promising approach could be to teach the pupils how to find and work with keywords on their own. This could also help pupils when they are engaging with English texts outside the classroom.

An intriguing relationship uncovered during the course of this quasi-experiment is the potentially reciprocal relationship between pupils' ability to infer word meaning and receptive vocabulary size. The results of this inquiry support previous research indicating a correlation between these variables. This aspect of contextual inferences becomes even more interesting when the effect of the contextual inferences score on the post-test keywords score are taken into consideration. The linear regression model made for group C indicates that this ability has had a large effect on the amount of new keywords that the pupils have learnt during the two teaching sessions (section 5.6). Moreover, Wilson (2013) demonstrated that this ability is trainable, unlike other individual differences related to word meaning such as the grammatical sensitivity component of the MLAT test battery (Skehan, 1989). In light of this, it is possible to argue strongly for the importance of teaching cognitive, and metacognitive strategies related to the use of contextual inferences in order to understand unknown vocabulary in the L2 classroom.

6. 4 Suggestions for further research

Because this thesis is based on only one study, with little support from previous research, several unanswered questions emerge. The main sections of this discussion have been oriented toward the theoretical and practical implications of this investigation. However, as was pointed out in section 6.1 this thesis is based on only one study, using a quasi-experiment research design with intact groups and a small sample size. In light of this, I have argued that the results are limited to only this setting, and cannot confidently be generalized to the population. Furthermore, the results of the pre-test have identified pre-existing differences between the experiment and control group. Thus, one important suggestion for further research is a call for new studies replicating, or improving on this quasi-experiment.

Because of the potential theoretical and practical implications of this study, I argue strongly for the importance of replicating research into the relationship between explicit focus on vocabulary and pupils' ability to infer word meaning using contextual cues. Preferably, these studies should use more than one class for each group condition, as well as random group assignment if possible. All testing instruments, except the vocabulary size test developed by Nation & Beglar (2007), used during this research project can be found in the appendix. If the results of new studies are in support of the hypotheses explored in this inquiry, the importance of contextual inferences and explicit focus on vocabulary will be

strengthened. This will further support my findings, and make generalizations to the population more possible.

Leaving aside the weaknesses and limitations of this study, several possible venues of further research can be proposed due to the results of this inquiry. These are related to the motivation of both pupils and teacher, and the possible interrelationship between contextual inferences, vocabulary, and other individual differences. Additionally, in light of the results of this study, the relationship between other teaching methods and contextual inferences are also a possible area of new research. The remainder of this discussion will be dedicated to these suggestions.

One important aspect that has not been explored during this research project is the effect of the two teaching methods on the motivation of both teacher and pupils in the two group conditions. The importance of motivation has been extensively documented. If the two different methods used during this quasi-experiment affect the motivation of pupils and teachers, it will have theoretical and practical implications. Because of the importance of self-efficacy and efficacy expectations (e.g., Skaalvik & Skaalvik, 2005; Reeves, 2009), one potentially significant effect of using explicit focus on vocabulary when dealing with difficult texts could be to positively affect the pupils' efficacy expectations and self-efficacy. The focus on keywords could in this instance be understood as a form of verbal persuasion (Reeve, 2009). In order to explore this aspect further, a mixed-methods approach using some form of qualitative research methods, such as interviews and group interviews, as well as surveys would be necessary.

Furthermore, both the pupils' and teacher's thoughts and attitudes towards using an unusual topic during the English sessions is another important dimension that has not been explored. This also includes the pupils' motivation for doing this form of activities in the L2 classroom. Arguably, these variables are highly important in determining the usefulness of this form of cross-curricular teaching. Thus, a predominately quantitative mixed-methods research design should be adopted in order to incorporate these aspects.

The second component of the pre-test has undergone a number of changes since the start of this project. Since its inception, there has always been a plan to investigate the effect of some form of individual difference on the effectiveness of the methodological approaches.

Because of the explicit focus on vocabulary, it was always planned to examine some form of individual difference which could possibly affect pupils' understanding of unknown words. Initially, it was planned to measure all three dimensions of language proficiency (FAC). However, this would need considerably more time than what was available for the pre-test. One alternative was to focus on pupils' accuracy, the hypothesis being that pupils who had a high level of accuracy would be better equipped to understand the grammatical function of the unknown word, which could then lead to a better understanding.

The idea of pupil accuracy having an effect on the usefulness of explicit focus on vocabulary soon lead to the concept of grammatical sensitivity. This would have been a highly interesting individual difference to examine as a variable in this quasi-experiment. However, because the MLAT tests are very difficult to gain access to, both in terms of cost and availability, it was not possible to use this as a variable in this thesis. As has been shown in section 3.7.2, there has also been quite a lot of criticism regarding the MLAT tests, and grammatical sensitivity in particular. The proposed fixedness or rigidity of grammatical sensitivity as an individual difference also means that it has limited practical value for teachers in a Norwegian context. One of the main purposes of the MLAT tests has always been to predict language learning success (lltf.net, 2011). This has limited value in the Norwegian primary and lower secondary schools, because all pupils must have English as a subject regardless of their supposed level of language aptitude.

It is important to note that this thesis has only explored the effect one individual difference on the effectiveness of using explicit focus on vocabulary. As was mentioned in section 3.7, there exists a multitude of individual differences, and all of these can potentially affect the results of using different methods and teaching activities (Skehan, 1989). I propose three such individual differences in relation to explicit focus on vocabulary and contextual inferences of unknown words. These individual differences are level of L2 proficiency, verbal ability, and grammatical sensitivity. To the best of my knowledge, these individual differences have not received attention from researchers, neither in relation to the use of explicit focus on vocabulary, or to explore if they are important components of pupils' ability to do contextual inferences. I postulate that these individual differences will be important components of pupils' ability to infer word meaning from context, and I argue for a study exploring the relationship between these variables. Furthermore, I propose that these

101

variables can be potentially important predictors for the usefulness of using an explicit focus on vocabulary. In light of this, new studies should incorporate these individual differences in the pre-tests, if possible.

CHAPTER 7 - CONCLUSION

This master's thesis has explored the use of teacher-initiated explicit focus on vocabulary in the form of keywords when working with a text about global warming and climate change (National Geographic, 2015). This was done using a two-group, interventionist, pre-test, post-test, delayed post-test research design in a Norwegian secondary school. This quasi-experiment has explored several aspects of this method.

The first aspect was to explore the usefulness of teacher-initiated explicit focus on vocabulary compared to an increased focus on content, by measuring pupils understanding of the main concepts in the text about global warming and their acquisition of new keywords. The results suggest that the explicit focus on vocabulary in the experiment group caused an increased amount of vocabulary item acquisition compared to the control group. Conversely, the focus on content lead to a greater amount of retention of the key concepts.

The second aspect was to determine the effectiveness of this method in addressing an individual difference between pupils. This individual difference was pupils' ability to infer the meaning of unknown words using contextual cues. The results indicate that the use of explicit focus on vocabulary mitigated the effect of this individual difference. Finally, the relationship between pupils' ability to infer word meaning from context and their vocabulary size was explored. The results found a strong relationship between these two individual differences.

Several weaknesses and limitations were identified and addressed. These were mostly related to the use of intact groups and small sample size. Furthermore, pre-existing differences between the groups related to receptive vocabulary size and previous knowledge of keywords were identified on the pre-test. In light of this, a major weakness of this study is that the results cannot be generalized to the population. Suggestions for further research were mainly related to replicating these results, preferably with larger samples, and random group assignment. Additionally, three other individual differences related to L2 language development were identified as possible components of pupils' ability to infer word meaning from contexts. These individual differences were proficiency, grammatical sensitivity, and verbal ability. Further research is needed to explore this potential relationship.

Nevertheless, the results of this thesis suggest that the use of explicit focus on vocabulary in the form of keywords is an effective way of helping pupils to learn new

103

vocabulary. Furthermore, it appears that the explicit focus on vocabulary in the experiment group has mitigated the effect of pupils' ability to infer word meaning from context. In light of these results, I make a strong case that time should be spent on explicit focus on vocabulary, and on improving pupils' ability to use contextual cues when encountering unfamiliar words in the L2 classroom.

BIBLIOGRAPHY

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: the interface between learning and assessment*. London: Continuum.

Atkinson, R. C. (1975). Mnemotechnics in second-language learning. *American Psychologist, 30*, 821–1. Retrieved from: http://dx.doi.org/10.1037/h0077029

Bandura, A. (1986). Self Efficacy. In Bandura, A. (Ed.) *Social foundations of though and action: A social cognitive theory* (pp. 390-453)*.* Englewood Cliffs NJ: Prentice Hall.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28,* 117–148.

Bandura, A. (1997). *Self-Efficacy: The exercise of control.* New York: W. H. Freeman.

Bauman, J. F., Edwards, C. E., Boland, E. M., Olejnik, S., Kame'enui, E. J. (2003). Vocabulary Tricks: Effects of Instruction in Morphology and Context on Fifth-Grade Students' Ability to Derive and Infer Word Meaning. *American Educational Research Journal Summer 2003, Vol. 40, No. 2, pp. 447-49.* Retrieved from: http://www.jstor.org/stable/3699395 21.04.2016

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, Mass.; London: MIT Press.

Bolger, D., J., Balass, M., Landen, E., & Perfetti, C., A. (2008). Context Variation and Definitions in Learning the Meanings of Words: An Instance-Based Learning Approach. *Discourse Processes, 45*(2), 122-159. Retrieved from: http:dx.doi.org/10.1080/01638530701792826

Brown, J. D., & Rodgers, T. S. (2002). *Doing second language research*. Oxford: Oxford University Press.

Carpenter, S. K., Sachs, R., E., Martin, B., Schmidt, K., & Looft, R. (2012). Learning new vocabulary in German: the effects of inferring word meanings, type of feedback, and time of test. *Psychon Bull Rev, 19*, 81-86. Retrieved from: doi: 10.3758/s13423-011-0185-7

Çetinavcı, B. M. (2013) Contextual factors in guessing word meaning from context in a foreign language. *Procedia - Social and Behavioral Sciences 116, 2670 – 2674.* Retrieved from: doi: 10.1016/j.sbspro.2014.01.633

Choo, L. B., Tan Ai Lin, D., & Pandian, A. (2012). Language Learning Approaches: A Review of Research on Explicit and Implicit Learning in Vocabulary Acquisition. *Procedia - Social and Behavioral Sciences, 55*, 852-860. Retrieved from: doi:10.1016/j.sbspro.2012.09.57

Chun, E., Choi, S., Kim, J. (2012). The effect of extensive reading and paired-associate learning on long-term vocabulary retention: An event-related potential study. *Neuroscience Letters, 521, 125-129.* Retrieved from: http://dx.doi.org/10.1016/j.neulet.2012.05.069

Coxhead, A., Nation, P., & Sim, D. (2015). Measuring the Vocabulary Size of Native Speakers of English in New Zealand Secondary Schools. *NZ J Educ Stud, 50*, 121-135. Retrieved from: http://doi:10.1007/s40841-015-0002-3

Coyle, D. (2010). *CLIL: Content and Language Integrated Learning* [Kindle]. Retrieved from: http://www.amazon.com/CLIL-Content-Language-Integrated-Learning-ebook/dp/B003VS0CM4/ref=sr_1_1?s=digital-text&ie=UTF8&qid=1461854114&sr=1-1&keywords=content+and+language+integrated+learning 28.04.16

Dancey, C. P. a., & Reidy, J. a. (2014). *Statistics without maths for psychology* (Sixth edition). Harlow: Pearson Education Limited.

de Leeuw, L., Segers, E., & Verhoeven, L. (2014). Context, Task, and Reader Effects in Children's Incidental Word Learning from Tex. *International Journal of Disability, Development and Education., 61*(3), 275-283. Retrieved from: http://dx.doi.org/10.1080/1034912X.2014.934010

Elley, W. B., & Manghubai, F. (1983). The Impact of Reading on Second Language Learning. *Reading Research Quarterly, 19*(1), 53-67. Retrieved from: http://www.jstor.org/stable/747337

Ellis, N., C. (1995). *The Psychology of Foreign Language Vocabulary Acquisition*. University of
      Wales Bangor. Retrieved from:

      https://pdfs.semanticscholar.org/da1b/4b40b75dd1293012aa685a89163c26997b63.
      pdf

Field, A. (2014). *Discovering Statistics using IBM SPSS Statistics* (4th edition). London: SAGE
      Publications Ltd.

Ghanbari, M., & Marzban, A. (2014). Effect of Extensive Reading on Incidental Vocabulary
      Retention. *Procedia - Social and Behavioral Sciences, 116*, 3854-3858. Retrieved from:
      doi:10.1016/j.sbspro.2014.01.854

Grigorenko, E. L., Sternberg, R. J., & Ehrman, M. E. (2000). A Theory-Based Approach to the
      Measurement of Foreign Language Learning Ability: The Canal-F Theory and Test. *The
      Modern Language Journal, 84*, 390-405. Retrieved from JSTOR.org website:
      http://www.jstor.org/stable/330568

Help Teaching. (2013). Context clues. Retrieved from:
      http://www.helpteaching.com/questions/Context_Clues 28.04.2016

Hannisdal & Jenset. (2012). *Analyzing English Grammar – Exercises for Advanced Students*.
      Bergen: Vigmonstad & Bjørke.

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency - Definitions,
      measurement and research. In *Dimensions of L2 performance and proficiency:
      Complexity, accuracy and fluency in SLA (2012)* (pp. 1-21). Amsterdam; Philadelphia:
      John Benjamins Pub. Co.

Knight, S. (1994). Dictionary Use While Reading: The Effects on Comprehension and
      Vocabulary Acquisition for Students of Different Verbal Abilities. *The Modern
      Language Journal, 78*(3), 285-299. Retrieved from:
      http://dx.doi.org/10.2307/330108

Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.

Larson-Hall, J., (2012) How to Run Statistical Analyses. In *Research methods in second
      language acquisition: a practical guide* [Kindle]. Retrieved from:

107

http://www.amazon.com/Research-Methods-Second-Language-Acquisition-ebook/dp/B006XNJCDE/ref=sr_1_1_twi_kin_2?s=books&ie=UTF8&qid=1461854225&sr=1-1&keywords=research+methods+in+second+language+acquisition

Larsen, R. J. & Buss, D. M. (2008) *Personality Psychology – Domains of Knowledge about Human Nature*. New York: McGraw-Hill.

Laufer, B. (1989). What percentages of text-lexis is essential for comprehension. In Lauren, C. & Nordman, M. (eds), *Special Language: From Humans Thinking to Thinking Machines.* Clevedon: Multilingual Matters.

Laufer, B., & Goldstein, Z. (2004). Testing Vocabulary Knowledge: Size, Strength, and Computer Adaptiveness. *Language Learning, 54*(3), 399-436. Retrieved from: DOI: 10.1111/j.0023-8333.2004.00260.x

Lightbown, P., & Spada, N. M. (2013). *How languages are learned* [Kindle]. Retrieved from:

http://www.amazon.com/Oxford-Handbooks-Language-Teachers-Languages-ebook/dp/B00BHHIYBC/ref=mt_kindle?_encoding=UTF8&me 28.04.2016

Lltf.net. (2011). The Modern Language Aptitude Test (MLAT). Retrieved from: http://lltf.net/wp-content/uploads/2011/09/Test_mlat.pdf

Loewen, L., & Philp, J. (2012) Instructed Second Language Acquisition. In Mackey, A., & Gass, S. M. (Ed) (2012). *Research methods in second language acquisition: a practical guide* [Kindle]. Retrieved from:

http://www.amazon.com/Research-Methods-Second-Language-Acquisition-ebook/dp/B006XNJCDE/ref=sr_1_1_twi_kin_2?s=books&ie=UTF8&qid=1461854225&sr=1-1&keywords=research+methods+in+second+language+acquisition 28.04.2016

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

Mackey, A., & Gass, S. M. (Ed) (2012). *Research methods in second language acquisition: a practical guide* [Kindle]. Retrieved from:

http://www.amazon.com/Research-Methods-Second-Language-Acquisition-ebook/dp/B006XNJCDE/ref=sr_1_1_twi_kin_2?s=books&ie=UTF8&qid=1461854225&sr=1-1&keywords=research+methods+in+second+language+acquisition

Matlin, M. W., & Matlin, M. W. C. (2009). *Cognitive psychology* (7th ed.), (International student Ed.). Hoboken: Wiley.

Nassaji, H. (2006). The Relationship between Depth of Vocabulary Knowledge and L2 Learners' Lexical Inferencing Strategy Use and Success. *The Modern Language Journal, 90(3),* 387-401*.* Retrieved from: http://www.jstor.org/stable/3876835 21.04.2016

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening*? Canadian Modern Language Review 63.1*, 59–82. Retrieved from:

http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/2006-How-large-a-vocab.pdf

Nation, I. S. P. (2007). The Four Strands. *Innovation in Language Learning and Teaching, 1*(1), 2-13. Retrieved from: doi:10.2167/illt039.0

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

http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Vocabulary-Size-Test-information-and-specifications.pdf 28.04.2016

National Geographic (2015). *Global warming.* Retrieved from:
http://education.**nation**algeographic.org/encyclopedia/global-warming/ 28.04.2016

Robinson, P. (2001b). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In Robinson, P. (Ed.). *Cognition and second language instruction (pp. 287-318)*. Cambridge: Cambridge University Press.

Robinson, P. J. (2002). *Individual differences and instructed language learning*. Philadelphia: John Benjamins Pub.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a Componential Framework for second language task design. *International Review for Applied Linguistics in Second Language Teaching (IRAL), 43(1), 1-32.*
109

Schatz E. K. & Baldwin, R. S. (1986) Context Clues Are Unreliable Predictors of Word Meanings. *Reading Research Quarterly, Vol. 21(4)*, 439-453. Retrieved from: http://www.jstor.org/stable/747615 21-04-2016

Reeve, J. M. (2009). *Understanding Motivation and Emotion.* Hoboken: John Wiley & Sons.

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*(04), 484-503. Retrieved from: doi:10.1017/S0261444812000018

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

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95(1),* 26–43. Retrieved from: http://www.jstor.org.galanga.hib.no/stable/41262309

Skehan, P. (1989). *Individual differences in second-language learning*. London; New York New York, NY: Edward Arnold; Distributed in the USA by Routledge, Chapman and Hall [Kindle]. Retrieved from:

http://www.amazon.com/Individual-Differences-Language-Second-Language-Acquisition/dp/0713166029/ref=sr_1_4?ie=UTF8&qid=1461853925&sr=8-4&keywords=individual+differences+in+second+language+acquisition

Skehan, P. (2002). Theorising and updating aptitude. In Robinson, P. J. (Ed.). *Individual Differences and Instructed Language Learning*. Philadelphia: John Benjamins Pub.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30(4), 510-532.*

Skaalvik, E. M. & Skaalvik, S. (2005). *Skolen som Læringsarena – Selvoppfatning, motivasjon og læring.* Oslo: Universitetsforlaget.

Shapiro, S. S. & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika, Vol. 52, No. 3/4,* 591-611. Retrieved from: http://sci2s.ugr.es/keel/pdf/algorithm/articulo/shapiro1965.pdf 28.04.2016

Sonbul, S., & Schmitt, N. (2010). Direct teaching of vocabulary after reading: is it worth the effort? *ELT Journal, 64*(3), 253-260. Retrieved from:

http://www.norbertschmitt.co.uk/uploads/sonbul-s-and-schmitt-n-(2010)-direct-teaching-of-vocabulary-after-reading-is-it-worth-the-effort-english-language-teaching-journal-64-3-253-260.pdf

Taber, K. (2007). *Classroom-based research and evidence-based practice: a guide for teachers*. London: SAGE.

UDIR. (2005a). *Core Curriculum for Primary, Secondary and Adult Education in Norway.* Retrieved from:

http://www.udir.no/globalassets/upload/larerplaner/generell_del/5/core_curriculum_english.pdf 28.04.016

UDIR. (2005b). *National Curriculum for Knowledge Promotion in Primary and Secondary Education and Training - The Quality Framework.* Retrieved from:

http://www.udir.no/globalassets/upload/larerplaner/fastsatte_lareplaner_for_kunnskapsloeftet/5/prinsipper_lk06_eng.pdf 28.04.2016

UDIR. (2013). *Primary and lower secondary education.* Retrieved from:
http://www.udir.no/Stottemeny/English/Curriculum-in-English/Curricula-in-English/

Vanpatten, B., Collopy, E., Josephe, P., Borst, S., & Qualin, A. (2013). Explicit Information, Grammatical Sensitivity, and the First-Noun Principle: A Cross-Linguistic Study in Processing Instruction. *The Modern Language Journal, 97*(2). doi:10.1111/j.1540-4781.2013.12007.x

Vanpatten, B., & Smith, M. (2015). Aptitude as Grammatical Sensitivity and the Initial Stages of Learning Japanese as a L2. *Studies in Second Language Acquisition, 37*, 135-165. Retrieved from: doi10.1017/S0272263114000345

Wilson, A. (2013). *The Impact of Inferring Word Meaning: Context Clues.* (Master's thesis), St. Catherine University. Retrieved from:
http://sophia.stkate.edu/cgi/viewcontent.cgi?article=1028&context=maed

28.04.2016

Zaid, M. A. (2009). A Comparison of Inferencing and Meaning-guessing of New Lexicon in Context versus Non-context Vocabulary Presentation. *The Reading Matrix, 9*(1). Retrieved from: http://www.readingmatrix.com/articles/zaid/article.pdf 28.04.2016

Appendix 1 - List of Keywords

- <u>Global</u> - relating to the whole world; worldwide:

- <u>Climate</u> - the weather conditions in an area in general or over a long period.

- <u>Average temperature</u> – the typical or normal temperature

- <u>Adapt</u> - become adjusted or used to new conditions

- <u>Absorb</u> - to take in a liquid, gas or other substance from the surface or space around

- <u>Atmosphere</u> - layers of gases surrounding a planet or other celestial body.

- <u>Degrees Celsius</u> – a temperature scale

- <u>Greenhouse effect</u> - phenomenon where gases allow sunlight to enter earth's atmosphere but make it difficult for heat to escape.

- <u>Insulating</u> - protect (something) by interposing material that prevents the loss of heat or the intrusion of sound.

- <u>Emission</u> – a discharge or release.

- <u>Natural range</u> – the normal, or typical amount seen or registered

- <u>Fossil fuels</u> - coal, oil, or natural gas. Fossil fuels formed from the remains of ancient plants and animals.

- <u>Diminishing</u> - to become smaller or less important.

- <u>Decaying</u> - becoming rotten or decomposing.

- <u>Glacier</u> - mass of ice that moves slowly over land.

- <u>Ancient</u> – very old

- <u>Precipitation</u> – rain or snow

- <u>Agricultural</u> - connected with the science or practice of farming

- <u>However</u> - used to introduce a statement that contrasts with something that has just been said

113

- <u>Sea level </u>- base level for measuring.

Appendix 2 – Parental consent form

Hei

Mitt navn er Christopher Loe Olsen og jeg er mastergradsstudent ved Høgskolen i Bergen. I forbindelse med min mastergrad har jeg planlagt å gjennomføre et prosjekt i samarbeid med [——], som er lærer i Engelsk i ditt barns klasse. Dette krever at du som forelder blir informert om, og gir samtykke til at ditt barn er med på dette prosjektet. I denne anledning sender jeg dette skrivet som gir en innføring i hva som skal gjennomføres slik at du som foresatt kan avgjøre om ditt barn skal være med på testingen i prosjektet. Det er viktig å merke seg at undervisningen vil være i tråd med de gjeldende læreplanene og testingen vil være helt anonym. Videre vil testene kunne brukes av [——] i forbindelse med hennes arbeid som lærer.

I forbindelse med dette prosjektet vil noe av klasseromsundervisningen i Engelsk brukes til å gjennomføre et forsøk hvor jeg vil se på effekten av to ulike undervisningsopplegg hvor målet er å bruke en tekst om global oppvarming og klimaendringer i engelskfaget. Det vil gjennomføres en test i forkant, og to tester i etterkant av undervisningen. Ingen personinformasjon vil bli lagret.

Dersom du ikke ønsker at ditt barn er med på testingen vil det være svært nyttig for meg at du krysser av under denne boksen slik av jeg vet at du har lest dette skrivet.

Med vennlig hilsen Christopher Loe Olsen

☐ Jeg har lest dette brevet, og samtykker til at mitt barn er med på prosjektet.

☐ Jeg har lest dette brevet, og ønsker ikke at mitt barn er med på prosjektet.

115

# Appendix 3 – Global warming and climate change text

Global Warming

Global warming describes the current rise in the average temperature of Earth's air and oceans. Global warming is often described as the most recent example of climate change.

Earth's climate has changed many times. Our planet has gone through multiple ice ages, in which ice sheets and glaciers covered large portions of the Earth. It has also gone through warm periods when temperatures were higher than they are today.

Past changes in Earth's temperature happened very slowly, over hundreds of thousands of years. However, the recent warming trend is happening much faster than it ever has. Natural cycles of warming and cooling are not enough to explain the amount of warming we have experienced in such a short time—only human activities can account for it. Scientists worry that the climate is changing faster than some living things can adapt to it.

In 1988, the World Meteorological Organization and the United Nations Environment Programme established a committee of climatologists, meteorologists, geographers, and other scientists from around the world. This Intergovernmental Panel on Climate Change (IPCC) includes thousands of scientists who review the most up-to-date research available related to global warming and climate change. The IPCC evaluates the risk of climate change caused by human activities.

116
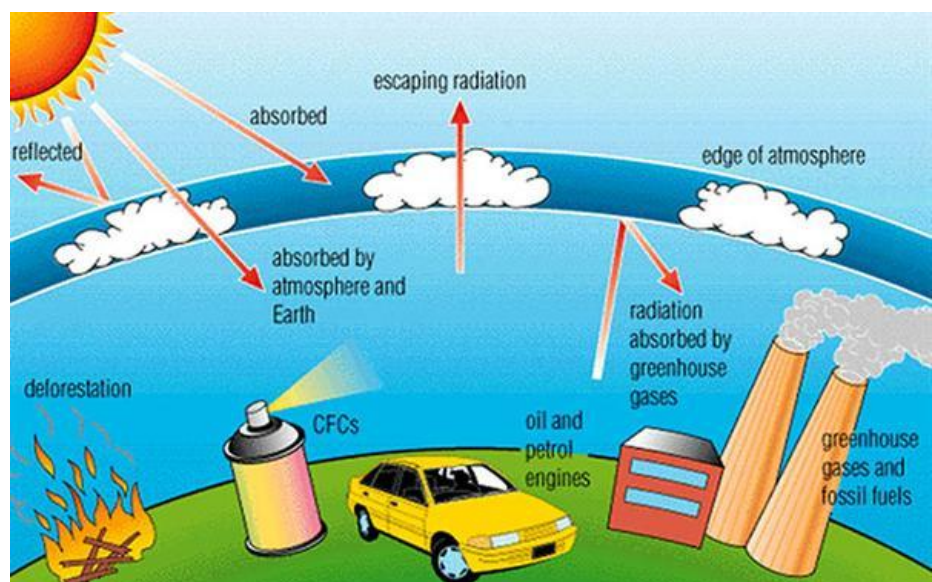
According to the IPCC's most recent report (in 2007), Earth's average surface temperatures have risen about 0.74 degrees Celsius (1.33 degrees Fahrenheit) during the past 100 years. The increase is greater in northern latitudes. The IPCC also found that land regions are warming faster than oceans. The IPCC states that most of the temperature increase since the mid-20th century is likely due to human activities.

The Greenhouse Effect

Human activities contribute to global warming by increasing the greenhouse effect. The greenhouse effect happens when certain gases—known as greenhouse gases—collect in Earth's atmosphere. These gases, which occur naturally in the atmosphere, include carbon dioxide, methane, nitrogen oxide, and fluorinated gases sometimes known as chlorofluorocarbons (CFCs).

Greenhouse gases let the sun's light shine onto the Earth's surface, but they trap the heat that reflects back up into the atmosphere. In this way, they act like the insulating glass walls of a greenhouse. The greenhouse effect keeps Earth's climate comfortable. Without it, surface temperatures would be cooler by about 33 degrees Celsius (60 degrees Fahrenheit), and many life forms would freeze.

Since the Industrial Revolution in the late 1700s and early 1800s, people have been releasing large quantities of greenhouse gases into the atmosphere. That amount has skyrocketed in the past century. Greenhouse gas emissions increased 70 percent between 1970 and 2004. Emissions of carbon dioxide, the most important greenhouse gas, rose by about 80 percent during that time. The amount of carbon dioxide in the atmosphere today far exceeds the natural range seen over the last 650,000 years.

Most of the carbon dioxide that people put into the atmosphere comes from burning fossil fuels such as oil, coal, and natural gas. Cars, trucks, trains, and planes all burn fossil fuels. Many electric power plants also burn fossil fuels. Another way people release carbon dioxide into the atmosphere is by cutting down forests. This happens for two reasons.

117

Decaying plant material, including trees, releases tons of carbon dioxide into the atmosphere. Living trees absorb carbon dioxide. By diminishing the number of trees to absorb carbon dioxide, the gas remains in the atmosphere.

Most methane in the atmosphere comes from livestock farming, landfills, and fossil fuel production such as coal mining and natural gas processing. Nitrous oxide comes from agricultural technology and fossil fuel burning.

All of these human activities add greenhouse gases to the atmosphere, trapping more heat than usual and contributing to global warming.

Effects of Global Warming

Even slight rises in average global temperatures can have huge effects. Perhaps the biggest, most obvious effect is that glaciers and ice caps melt faster than usual. The meltwater drains into the oceans, causing sea levels to rise and oceans to become less salty.

Ice sheets and glaciers advance and retreat naturally. As Earth's temperature has changed, the ice sheets have grown and shrunk, and sea levels have fallen and risen. Ancient corals found on land in Florida, Bermuda, and the Bahamas show that the sea level must have been 5 to 6 meters (16-20 feet) higher 130,000 years ago than it is today. Earth doesn't need to become oven-hot to melt the glaciers. Northern summers were just 3 to 5 degrees Celsius (5-9 degrees Fahrenheit) warmer during the time of those ancient fossils than they are today.

However, the speed at which global warming is taking place is unprecedented. The effects are unknown.

Many scientists use the term "climate change" instead of "global warming." This is because greenhouse gas emissions affect more than just temperature. Another effect involves changes in precipitation like rain and snow. Patterns in precipitation may change or become more extreme. Over the course of the 20th century, precipitation increased in eastern parts

of North and South America, northern Europe, and northern and central Asia. However, it has decreased in parts of Africa, the Mediterranean, and parts of southern Asia.

Appendix 4 – Testing instruments
Pre-test component 1

Her skal du lese setningene og så forsøke å forstå hva de understrekede ordene betyr ut i fra det som står i setningen. Skriv hva du tror kan være et annet ord for det understrekede ordet i den blanke boksen. Dersom du ikke klarer å skrive et engelsk ord, kan du bruke norsk. Husk at du har begrenset med tid, så ikke vent for lenge før du går over til det neste ordet! ☺

Sentence:	Meaning:
Instead of having friends among the other animals, the rabbit has many <u>FOES.</u>	Enemies. (om du ikke kommer på et engelsk ord kan du skrive på norsk. For eksempel <u>fiender</u>)
"That roller coaster was a <u>THRILL</u>!" Chris shouted, eyes wide open and a grin on his face.	
They looked <u>PELTIVELY</u> for the hole.	
Jeff's father was <u>IRATE</u> because Jeff talked back to him and had a bad attitude all day long.	
When Chris heard the doorbell ring, she <u>TORE</u> down the stairs so quickly that she almost tripped.	
The customers were <u>SCONZIATIC</u> over the lack of service!	
Molly has so many friends because she is a <u>GREGARIOUS</u> person.	

The colour of the fence was getting faint so we <u>BLAYNED</u> it.	
Tuesday was Moira's first day in the new school. She was confused by the endless hallways and was often lost. By Friday, however, her <u>BEWILDERMENT</u> had ended	
Thomas hiked to the APEX of the mountain and got a great view of the entire city.	

Compontent 2

Pre-test Section Three:

Her skal dere krysse av på linjen ved siden av den definisjonen som passer best til ordet som er uthevet. Det er lov å gjette!

1) **Global:**

__ Relating to the whole world; worldwide.

__ Next to or near something.

__ An object shaped like a ball with a map of the world on its surface, usually on a stand so that it can be turned.

__ Of very great excellence or beauty

2) **Unprecedented**

__ In spite of that; notwithstanding; all the same.

__ Not connected with the position or activities of a president.

__ Never before known or experienced.

3) **Decaying**

__ A period of ten years.

__ Becoming rotten or decomposing.

__ Counting Something.

__ Measuring a distance.

4) **Emission**

__ A discharge or release.

__ A long search for something.

__ A compound found in living tissue.

__ A letter or message containing information or news.

5) **Climate**

__ The weather conditions prevailing in an area in general or over a long period.

__ The areas of land covered by shallow water or saturated by water.

__ Necessary or very important.

__ Severe weather indicating a disturbed state of the atmosphere resulting from uplifted air.

6) **Precipitation**

__ Drops of liquid that appear on the surface of your skin when you are hot, ill/sick or afraid.

__ To take in a liquid, gas or other substance from the surface or space around.

__ Rain or Snow

__ Full of power to refresh refreshing.

7) **Agricultural**

__ Necessary or very important.

__ Connected with the science or practice of farming.

__ Related to the production of fuel.

___ Part of the shipping industry.

8) **Greenhouse**

___ A shop which sells vegetables.

___ Someone who does not have a lot of work experience.

___ Building, often made of glass or other clear material, used to help plants grow.

___ A tropical bird species.

9) **Atmosphere**

___ Layers of gases surrounding a planet or other celestial body.

___ The deepest area of the ocean.

___ Another word for submarine.

___ An important part of an airplane.

10) **Glacier**

___ An inland sea.

___ Mass of ice that moves slowly over land.

___ The frosting on a cake.

___ Some who works with making glass.

11) **Fossil Fuels**

___ Fuels consisting of either coal, oil, or natural gas.

___ Fuels consisting of either carbon Monoxide, Carbon Dioxide, or Chloride.

___ Fuels consisting of either biofuel, Hydrogen, or Alcohol.

___ Fuels consisting of either plutonium, uranium, or caesium.

12) **Insulating**

123

__ Making fun of someone.

__ Covering something with earth in order to keep it dry.

__ Protect (something) by interposing material that prevents the loss of heat.

__ The act of wrapping up a present.

Post-test component 1

Her skal dere krysse av på linjen ved siden av den definisjonen som passer best til ordet som er uthevet, akkurat slik som sist ☺ . Det er lov å gjette!

1) **Unprecedented**

__ In spite of that; notwithstanding; all the same.

__ Not connected with the position or activities of a president.

__ Never before known or experienced.

2) **Global:**

__ Relating to the whole world; worldwide.

__ Next to or near something.

__ An object shaped like a ball with a map of the world on its surface, usually on a stand so that it can be turned.

__ Of very great excellence or beauty

3) **Climate**

__ The weather conditions prevailing in an area in general or over a long period.

__ The areas of land covered by shallow water or saturated by water.

__ Necessary or very important.

__ Severe weather indicating a disturbed state of the atmosphere resulting from uplifted air.

4) **Precipitation**

124

__ Drops of liquid that appear on the surface of your skin when you are hot, ill/sick or afraid.

__ To take in a liquid, gas or other substance from the surface or space around.

__ Rain or Snow

__ Full of power to refresh refreshing.

5) **Emission**

__ A discharge or release.

__ A long search for something.

__ A compound found in living tissue.

__ A letter or message containing information or news.

6) **Insulating**

__ Making fun of someone.

__ Covering something with earth in order to keep it dry.

__ Protect (something) by interposing material that prevents the loss of heat.

__ The act of wrapping up a present.

7) **Agricultural**

__ Necessary or very important.

__ Connected with the science or practice of farming.

__ Related to the production of fuel.

__ Part of the shipping industry.

8) **Atmosphere**

__ Layers of gases surrounding a planet or other celestial body.

__ The deepest area of the ocean.

__ Another word for submarine.

__ An important part of an airplane.

9) **Glacier**

__ An inland sea.

__ Mass of ice that moves slowly over land.

__ The frosting on a cake.

__ Some who works with making glass.

10) **Fossil Fuels**

__ Fuels consisting of either coal, oil, or natural gas.

__ Fuels consisting of either carbon Monoxide, Carbon Dioxide, or Chloride.

__ Fuels consisting of either biofuel, Hydrogen, or Alcohol.

__ Fuels consisting of either plutonium, uranium, or caesium.

11) **Greenhouse**

__ A shop which sells vegetables.

__ Someone who does not have a lot of work experience.

__ Building, often made of glass or other clear material, used to help plants grow.

__ A tropical bird species.

12) **Decaying**

__ A period of ten years.

__ Becoming rotten or decomposing.

__ Counting Something.

__ Measuring a distance.

Post-test Component 2

Dette er en ny flervalgsoppgave. Her skal dere svare på spørsmålet ved å sette ring rundt bokstaven før det mest rette svaret ☺

1) How has global warming happened before?

    a) Global warming has happened once before, after the last ice age.

    b) Global warming has never happened before, it is completely man-made.

    c) Global warming has happened before, it occurs every 50 000 years.

    d) Global warming has happened before, it has gone through multiple ice ages and warm periods.

2) How do scientists believe that global warming is being different from how it has happened before?

    a) It is no longer happening every 50 000 years.

    b) There has not been an ice age since the last warm period.

    c) It is occurring much faster than what has happened before.

    d) It is only happening in the northern hemisphere.

3) How does the greenhouse effect happen?

    a) The greenhouse effect happens by the earth's atmosphere allowing the sun's radiation to enter and traps the heat.

    b) The greenhouse effect happens by trapping the heat that is created by the earth's volcanoes.

    c) The greenhouse effect happens by not allowing the trees to cool down the earth.

    d) The greenhouse effect happens by trapping all of the sun's heat in the oceans.

4) What would happen on earth without the greenhouse effect?

    a) There would be no global warming, and the sea levels would not rise.

127

b) The earth's temperature would be much cooler, which would save most of the lifeforms on earth.

c) The earth's temperature would be much cooler, and many lifeforms would freeze.

d) There would be no air left on earth, and all life would die.

5) How has the earth's average temperature changed the last 100 years?

a) The earth's average temperature has risen about 0.74 degrees Celsius the last 100 years.

b) The earth's average temperature has risen about 15 degrees Celsius the last 100 years.

c) The earth's average temperature has risen about 5 degrees Celsius in the last 100 years.

d) The earth's average temperature has not risen in the last 100 years.

6) Why does many scientists use the term *climate change* instead of *global warming*?

a) Because the earth's average temperature is not actually rising.

b) Because global warming is only happening in the northern hemisphere.

c) Because global warming is not only affecting the temperature but also the weather.

d) Because it is happening to quickly to call it global warming.

7) What are the most common greenhouse gasses?

a) Coal, oil, and plastic.

b) Oxygen, hydrogen, and xenon.

c) Carbon dioxide, methane, and CFCs.

d) Agriculture, fluoride, and fossil fuels.

8) What is Global warming?

a) Global warming is the rise of the earth's average temperature.

128

b) Global warming is the rise of the winter temperature in the northern hemisphere.

c) Global warming is the appliance of heat into a sphere shaped object.

d) Global Warming is the rise in the earth's temperature which happens every summer.

9) Where does most of the methane in the atmosphere come from?

a) From cars and refrigerators.

b) From agriculture and fossil fuel production.

c) From aerosol cans and foresting.

d) From the fishing industry and shipping.

10) What is the IPCC?

a) The IPCC is a group of politicians who are in charge of the UN climate panel.

b) The IPPC is a group of scientists who are working with the latest research into global warming and climate change.

c) The IPCC is a group of politicians who do not believe in climate change.

d) The IPCC is a group of activists who work with raising peoples awareness of global warming and climate change.

Section 3

Please explain in English how the earth's atmosphere is similar to the glass walls of a greenhouse.

Write inside the box! ☺

Appendix 5 – Teaching plans

Teaching Plan Session One – Experiment Group (45 minutes)

1-5 minutes – Used to start up the session. The pupils find their English writing book and the teacher hands out the global warming and climate change text. No other books or teaching materials are needed.

5-20 minutes – The teacher leads a PowerPoint presentation, presenting keywords and the subject text.

20-40 minutes – The pupils work with the subject text, a PowerPoint slide with the keywords is present. They read through it underlining the keywords and focusing on other difficult words, which they write down. Those who are finished early will start on their homework.

40-45 minutes – Wrapping up the session. The pupils bring their English writing book and the subject text home, and are instructed NOT to lose it!

Goals of this Session:

This session is meant to familiarize the pupils with the subject text and the keywords. After this session, it is expected that the pupils are familiar with a number of the new keywords, specifically that they are able to recognize them and their meaning in context.

Knowledge Promotion Competence Aims:

"Understand the main content and details of texts one has chosen"

"Read, understand and evaluate different types of texts of varying length about different topics"

"Communicate and converse about contemporary and academic topics"

Teaching Plan Session One – Control Group

1-5 minutes – Used to start up the session. The pupils find their English writing book and the teacher hands out the global warming and climate change text. No other books or teaching materials are needed.

5-20 minutes – The teacher leads a PowerPoint presentation, presenting the subject text and some of the key concepts.

20-40 minutes – The pupils work with the subject text. On the PowerPoint there are a number of content specific questions that they are meant to find the answer to. Those who are finished early will start on their homework.

40-45 minutes – Wrapping up the session. The pupils bring their English writing book and the subject text home, and are instructed NOT to lose it!

<u>Goals of this session</u>

The goals of this session are to familiarize the pupils with the use of the keywords and how they are related to the subject specific text. After the session it is expected that the pupils know how to use some of the keywords and how these are related to the subject.

<u>Knowledge Promotion Competence Aims:</u>

"Understand the main content and details of texts one has chosen"

"Read, understand and evaluate different types of texts of varying length about different topics"

"Communicate and converse about contemporary and academic topics".

Teaching Plan Session Two – Experiment Group

1-5 minutes - Used to start up the session. The pupils find their English writing book and the teacher hands out the global warming and climate change text. No other books or teaching materials are needed.

5-20 minutes – Teacher lead PowerPoint presentation. In this presentation the teacher leads a class discussion of the difficult words that the pupils have found during their homework. Afterwards some of the keywords are presented and explained *in context* using examples from the subject text.

20-40 minutes – The pupils work with two keyword specific tasks related to the subject text. One is a fill in the missing word task, and the other is a productive task where the pupils construct their own subject specific sentences using the keywords and subject text as inspiration.

131

40-45 minutest – Wrapping up the session. The pupils put the subject text in their English writing book.

Goals of this session

The goals of this session is to familiarize the pupils with the keywords and how they are related to the subject specific text. After the session it is expected that the pupils know how to use some of the keywords and how these are related to the subject.

Knowledge Promotion Competence Aims

 "Understand the main content and details of texts one has chosen"

 "Read, understand and evaluate different types of texts of varying length about different topics"

"Communicate and converse about contemporary and academic topics"

Teaching Plan Session Two – Control Group

1-5 minutes - Used to start up the session. The pupils find their English writing book and the teacher hands out the global warming and climate change text. No other books or teaching materials are needed.

5-20 minutes – Teacher lead PowerPoint presentation. In this presentation the teacher leads a class discussion of the subject specific factual sentences that the pupils have produced during their homework. Afterwards some more of the main concepts from the subject text are presented.

20-40 minutes – The pupils work with two content specific tasks related to the subject text. One is a true or false task using statements from the subject specific text, and the other is a productive task where the pupils write a short review or summary of the text.

40-45 minutest – Wrapping up the session. The pupils put the subject text in their English writing book.

Goals of this session

The goals of this session are to familiarize the pupils with more of the concepts of the subject text. After the session, it is expected that the pupils have a good understanding of what global warming is, what causes it, and its possible consequences.

Knowledge Promotion Competence Aims

 "Understand the main content and details of texts one has chosen"

"Read, understand and evaluate different types of texts of varying length about different topics"

"Communicate and converse about contemporary and academic topics"

Appendix 6 – Fill in missing word task

<u>Fill in the Missing Words</u>

In this task you are going to fill in the missing word in the sentences. Each sentence has gotten one or more words removed and replaced with a line_____ you are going to write the missing word in your rough book, using the text, and the <u>keywords sheet</u>! All of the sentences can be found in the text! ☺

1) Earth's _____ has changed many times.

2) _____, the recent warming trend is happening much faster than it ever has.

3) Nitrous oxide comes from _____ technology and fossil fuel burning.

4) These gases, which occur naturally in the _____, include carbon dioxide, _____, nitrogen oxide, and fluorinated gases sometimes known as chlorofluorocarbons (CFCs).

5) Many scientists use the term "_____" instead of "global warming."

6) Patterns in_____ may change or become more extreme.

7) Scientists worry that the climate is changing faster than some living things can _____ to it.

8) _____ of carbon dioxide, the most important , _____ rose by about 80 percent during that time.

9) Most of the carbon dioxide that people put into the atmosphere comes from burning _____ such as oil, coal, and natural gas.

10) Most methane in the _____ comes from livestock farming, landfills, and _____ production such as coal mining and natural gas processing.

Appendix 7 – True or false task

<u>True OR False</u>

This is a "true or false task". I have made a number of statements. Some of these are true, and some are false. Your task is to identify and write down if the statement it true or false. Use the text to find out if the statements are true of false! ☺

1) "Global warming has never happened before". True or False? _____

2) "The speed at which global warming is taking place is unprecedented".

True or False? _____

3) "Past changes in Earth's temperature happened very slowly, over hundreds of thousands of years."

True or False? _____

4) "The Earth has never been this warm before".

True or False? _____

5) "Our planet has gone through only one ice age" True or False? _____

136

6) "Earth's average surface temperatures have risen about 25 degrees Celsius (30 degrees Fahrenheit) during the past 100 years.

True or False? _____

7) "Human activities do not contribute to global warming by increasing the greenhouse effect."

True or False? _____

8) "Scientists think that the climate is changing so slowly that living things can adapt to it."

True or False? _____

9) "The earth has also gone through warm periods when temperatures were higher than they   are   today".

True or False? _____

10) "The IPCC is a team of scientist who do not believe in human made climate change".

True or False? _____

Appendix 8 – Observation Schedule

Section One – Noise Level

Time	Noise Level	Notes
10 min		
20 min		
30 min		
40 min		
50 min		
60 min		

Section Two – Deviation From Schedule

Time	Notes

138

Section 3 – Deviation From Teaching Plan

Time	Notes

Appendix 9 – Teacher debrief questionnaire

Questions For Debrief:

How do you feel this session went?

To which extent do you feel that you followed the time schedule from the teaching plan?

To which extent do you feel that you followed the teaching plan?

How do you feel that using the Power Points went?

How do you feel the pupils responded to the teaching activities in this session?

140