Reading comprehension in the second language

The relationship between second language reading comprehension, second language vocabulary and first language reading comprehension

Areti Kalaitzi

Special Needs Education

40

University of Oslo

Faculty of Educational Sciences

Spring 2020

Reading comprehension in the second language

The relationship between second language reading comprehension, second language vocabulary and first language reading comprehension in adult native Norwegian speakers.

2020

The relationship between second language reading comprehension, second language vocabulary and first language reading comprehension in adult native Norwegian speakers.

Areti Kalaitzi

http://www.duo.uio.no

Print: University of Oslo

Abstract

Background and research question

Reading comprehension in one's first language is a broadly researched phenomenon. However, technological, social and economic changes have led to an increase of people learning to read in a foreign language (Saiegh-Hadad & Geva, 2010; Bernhardt, 2011). Thus, there is an increasing interest in research about second language reading. Reading in the second language is similar to, but at the same time, different in many ways than reading in the first language (Jiang, 2011; Yamashita & Shiotsu, 2017). It is important to understand which skills play a significant role in the development of second language reading. It is well established by existing research that second language vocabulary knowledge and first language reading ability are important components of second language reading comprehension (Raudszus et al., 2018; Raudszus et al., 2019; Jeon & Yamashita, 2014; Yamashita & Shiotsu, 2017; O'Connor et al., 2019; Schaars et al., 2019; Verhoeven, Voeten & Vermeer, 2019; Brevik et al., 2016). Furthermore, cross-linguistic transfer of skills has been strongly suggested between the first and the second language (Verhoeven, Perfetti & Pugh, 2019). Some theories have been suggested about the interaction and the transfer between the two languages (Cummins, 1979; Clarke, 1978; Bernhardt, 2011). However, there is not yet an agreement about how and when this transfer occurs. Moreover, there is lack of studies on the transfer of knowledge between the two languages with variation in the language background (Raudszus et al., 2018; Park, 2013). This study's purpose is to investigate the relationship between second language reading comprehension, second language vocabulary knowledge and first language reading comprehension with Norwegian as the first and English as the second language. It is also desired to see whether there is linguistic transfer of knowledge between Norwegian and English. The study is based on the following research question: *To what extent does L2 (English) vocabulary account for L2 reading comprehension, after accounting for L1 (Norwegian) reading comprehension, in adult native Norwegian speakers?*

Method

The study is based on data collected for the Norwegian part of an international study named Multilingual Eye-Tracking Corpus (MECO) (Cop et al., 2017) about eye-movements of adults during reading in many languages with different writing systems. The sample consists

of 50 young adults with Norwegian as their first language. A quantitative approach has been used and the participants have been tested with various psychometric tests. The study is non-experimental and no variables were manipulated. Multiple hierarchical regression analysis has been used to investigate to what extend first language reading comprehension and second language vocabulary knowledge can predict second language reading comprehension in adult native Norwegian speakers. In addition, bivariate correlation analysis has been conducted to investigate the relationship between these variables.

Results

In line with previous research, the correlation analysis showed that both first language reading comprehension and second language vocabulary knowledge correlate strongly with second language reading comprehension. The hierarchical regression analysis showed that first language reading comprehension and second language vocabulary accounted for the half of second language reading comprehension's variance. This finding is in line with the existing theory about cross-linguistic transfer and the factors that influences it.

Acknowledgements

First, I would like to thank Athanasios Protopapas for the great supervision. Thank you for all the constructive conversations, the educational and motivating feedback and for always taking time to answer any question and make sure everything is clear.

I would also like to thank Laoura Ziaka for her patience and thoroughness in our training on using the eye-tracking device, for always taking time to fix any problem and for making the friendliest possible environment for us to train and collect data.

Special thanks to Kelly Nisbet for traveling from Canada to help us with training and preparation for data collection and with relevant literature.

Also, I would like to thank Vasiliki Diamanti and Anders Lunde for helping us with some of the Norwegian tests.

Thanks to my great fellow students, Sara Maria Da Fonseca and Veronica Tønnesen for being an awesome team and for helping and supporting each other through all the way. Without your help and support this project could not be done.

Thanks to all the people that took part in our research for taking the time and energy to come and participate.

Last, but not least, I would like to thank Alexandros Kentros for fixing some important technical problems and for always supporting and encouraging me.

Table of Contents

1. Introduction

1.1 Background and purpose

In today's world reading is a crucial skill. People need reading in order to survive, to get informed, to communicate, to work etc. It plays an important role in our social and working lives (Cain, 2010; Castles et al., 2018; Li & Clariana, 2019; Raudszus et al., 2019). As expected, how children learn to read has been in focus of research for many decades, followed by arguments of how children should be taught to read (Castles et al., 2018). The goal of reading is to understand, so reading comprehension is the goal of reading. Identifying the skills involved in reading comprehension will allow us to recognize signs of difficulties in reading comprehension and provide us with the needed information about how reading comprehension should be built and developed (Hjetland et al., 2017). Studying how adult skilled readers process reading comprehension, will lead to a better understanding of what beginning readers need to learn. Skilled readers tend to process written material almost automatically in many cases and can easily retrieve the meanings of words, generate sentence meaning and link them successfully in a coherent representation of the text's meaning (Cain, 2010). Thus, understanding what skills and processes adult readers use for reading comprehension, will enable the identification of the skills that children need to be taught for reading development.

In the developed world it is very easy for people to move around and there is an enormous amount of available information in many different languages on the internet. In addition, unstable political and economic situations in many countries have led people to immigrate. Thus, there are many, both children and adults, that learn to read in another language than their mother tongue (Verhoeven, Voeten & Vermeer, 2019; Raudszus et al., 2018; Saiegh-Hadad & Geva, 2010; Bernhardt, 2011). Their first language is acquired in preschool years by being exposed to it at home but they have to acquire the second language later in their life and this could be a challenging task (Verhoeven, Voeten & Vermeer, 2019). As the need for learning to read and comprehend in a second language increases, the need to identify what processes and skills are involved and support reading comprehension in a second language becomes bigger.

Vocabulary in the target language has been found to be a very important predictor of reading comprehension in both the first and the second language (Castles et al., 2018; Bernhardt, 2011; Chung et al., 2019; Brevik et al., 2016; Verhoeven, Voeten & Vermeer, 2019;

Raudszus et al., 2018; Jeon & Yamashita, 2014; Yamashita & Shiotsu, 2017; O'Connor et al., 2019; Schaars et al., 2019). Furthermore, first language vocabulary has also been found to have a positive effect on second language reading comprehension in a few studies (Schaars et al., 2019; Chung et al., 2019) as well as first language reading ability (Brevik et al., 2016). In fact, there is found to be a transfer of acquired skills and processes between the first and the second language (Chung et al., 2019). However, this transfer depends on many factors that concern similarities and differences between the two languages and the level of the reader's language knowledge (Chung et al., 2019).

In the last few decades there has been a significant progress in understanding the mechanisms between single word reading in typical adult native speakers but there is limited research concerning sentence-level and text-level reading comprehension, especially in someone's second language (Li & Clariana, 2019; Chung et al., 2019). In addition, studies concerning the role of first language knowledge in second language reading comprehension are sparse and their results are mixed (Raudszus et al., 2018). Raudszus et al. (2018) and Park (2013) also point out that there should be more future studies on the transfer of knowledge between the two languages with variation in the language background. English as a second language has been widely studied whereas for Norwegian as the first language the research is still limited. Studies based on this combination of languages will allow for a better understanding of the transfer between them and will enrich the pool of available information for comparison of the transfer between different first languages and English as a second language.

The research problem being investigated is:

To what extent does L2 (English) vocabulary account for L2 reading comprehension, after accounting for L1 (Norwegian) reading comprehension, in adult native Norwegian speakers?

1.2 Structure of the study

In the introduction the role of second language vocabulary and first language reading ability in second language reading comprehension has been justified as the purpose of the study. In the following chapters there will be a deepening in the topic by presenting the various components of reading comprehension in both first and second language. To understand the development and function of reading and reading comprehension, models about reading and

reading comprehension will be presented. Then there will be a presentation of the leading hypotheses about how transfer between the first and the second language occurs.

In the third chapter the methodological process of the study will be presented. Design, participants, variables and statistical analyses will be described along with validity, reliability and ethics. In chapter 4 the statistical analysis and its results will be presented, while in chapter 5 there will be a discussion of the results in light of the theory and data presented earlier.

2. Theoretical background

2.1 Reading

Nowadays reading is a critical skill (Li & Clariana, 2019). People that live in the developed world cannot imagine their everyday life without print and written text. We use reading in every aspect of our life from driving and cooking to learning and working (Raudszus et al., 2019). It enables people to access information about the essentials of their everyday life, follow instructions, navigate through unknown places, communicate with each other, acquire and share knowledge, find work opportunities etc. (Cain, 2010). Reading can transform lives (Castles et al., 2018). But reading is not just a single skill. It is a combination of a range of knowledge, abilities and strategies (Cain, 2010).

When we read the goal is to understand what was written (Castles et al., 2018). When reading, the reader needs to access the meaning of every individual word (Cain, 2010). The definition of reading according to Rayner et al. (2016) is to process the information of a text in order to recover the meaning of every word, phrase and sentence. In line with this is the definition of reading literacy provided by PISA (OECD, 2006 in Bernhardt, 2011 p.16-17) which suggests that "reading literacy is understanding, using and reflecting on written texts, in order to achieve one's goals, to develop one's knowledge and potential and to participate in society". This definition implies that reading literacy is more than decoding and literal comprehension. It suggests that the reader has both an active and interactive role in gaining meaning of what is written (Bernhardt, 2011). Thus, to read successfully more than recognizing a sequence of individual words is needed. The reader also needs to understand the relationship between them and make inferences about what was not written (Rayner et al., 2016).

As mentioned above, reading is depending on writing, we only read what is written. Writing is composed by marks and fine lines and therefore the acuity limits of vision constrain the reading process. Human visual acuity is limited and that is the reason why we move our eyes when we read. These eye movements are very small and can be tracked by eye-tracking devices (Rayner et al., 2016). These devices are computing the location of the eye with a quite high precision, up to a thousand times per second. This allows us to know at which word and exactly where in the word the reader is looking with precision to the millisecond so we can then separate these information and find the exact time that the eyes remain in the same location (fixation) and the time they move from one location to the next (saccade).

Another information we can get from the movements of the eyes while reading is how many and which words the reader tends not to look at but move to the next one (skips) and when the reader is looking again at a prior word (regression). These skips often happen with words that are very short, very frequent or easily predicted from the context. Regressions often happen due to failure in comprehension (Holmqvist et al., 2011).

Reading is not an easy activity. It involves a lot of different skills, processes and types of knowledge (Cain, 2010). The reader uses his/her semantic, syntactic and graphic knowledge to gain the meaning of every word (Castles et al., 2018). Huey (1968 in Cain, 2010 p.3) points out that "to completely analyze what we do when we read, would almost be the acme of the psychologist's achievements, for it would be to describe many of the most intricate workings of the human mind". Thus, there is need to break the activity of reading into smaller chunks that would be easier to analyze and understand. Cain (2010) following the Simple View of Reading (Gough and Tunmer, 1986) suggests that reading should be divided in two components, word reading and linguistic comprehension.

2.1.1 The Simple View of Reading

According to the Simple View of Reading (SVR) (Gough & Tunmer, 1986) reading is a product of word decoding and linguistic comprehension showed by the following relationship, which looks like an equation, $R = D \times C$ (Gough & Tunmer, 1986; Cain, 2010; Yamashita & Shiotsu, 2017; Lervåg & Grøser Aukrust, 2010; van den Bosch et al., 2019; Yaghoub Zadeh et al., 2012; Hjetland et al., 2017; Castles et al., 2018). The R term represents reading, the D term represents decoding and the C term represents linguistic comprehension (Cain, 2010). Decoding is the mapping of print to sound, the efficient word recognition and linguistic comprehension refers to the processes involved in understanding words and the sentences and discourse they produce (Gough & Tunmer, 1986; Cain, 2010; Lervåg & Grøser Aukrust, 2010; Hjetland et al., 2017). Linguistic comprehension is usually measured with and operationally equated to listening comprehension (Gough & Tunmer, 1986; Cain, 2010; Yamashita & Shiotsu, 2017; Verhoeven, Voeten & Vermeer, 2019). Research suggests that the product of decoding and linguistic comprehension is not just reading but reading comprehension, since the aim of reading is to understand what is written (Cain, 2010; Castles et al., 2018; Lervåg & Grøser Aukrust; van den Bosch et al., 2019; Yaghoub Zadeh et al., 2012; Hjetland et al., 2017).

The difference between SVR and earlier reading models is that SVR proposes that reading is a product of decoding and linguistic comprehension, whereas earlier models proposed that it is the sum of component skills (Cain, 2010). This means that if one of the component skills is nil, then the overall reading ability will be zero. Neither decoding nor linguistic comprehension is sufficient alone. If a reader cannot decode words, but can understand spoken language, reading ability will be zero. Again, if one knows how to decode words but has no understanding of the spoken language, the reading ability will be zero (Gough & Tunmer, 1986; Cain, 2010).

The Simple View of Reading suggests that the relations between reading ability, word decoding and linguistic comprehension will change over time as the reader becomes more skilled (Cain, 2010). The contribution of word decoding is larger in beginning readers, whereas the contribution of linguistic comprehension is larger on more skilled readers (Yamashita & Shiotsu, 2017). For beginning readers, the major difficulty is to decode words, therefore reading ability is limited by decoding difficulties. When the reader masters this skill and decoding becomes more automatized, linguistic comprehension will be a more important factor of reading ability. Thus, as reading ability progresses in development, linguistic comprehension becomes a more important factor for successful reading than decoding (Cain, 2010; Yamashita & Shiotsu, 2017; Lervåg & Grøser Aukrust, 2010; Hjetland et al., 2017).

However, despite the broad appeal of the Simple View of Reading, it is not sufficiently specified (Yaghoub Zadeh et al., 2012). First, it is a framework, not a model which means that it does not explain how decoding and linguistic comprehension develop or operate (Castles et al., 2018). Second, there is inconsistency in how the key components are defined and measured (Cain, 2010; Castles et al., 2018). Especially for linguistic comprehension definitions and measures range from vocabulary and storytelling to grammar, inference making and verbal short-term memory (Cain, 2010; Castles et al., 2018; Hjetland et al., 2017). Finally, it does not include the role of reading fluency in decoding and the role of underlying cognitive processes in comprehension (Cain, 2010; Yaghoub Zadeh et al., 2012).

2.1.2 Skilled reading

It is important to examine how skilled readers process words and understand text in order to study reading development. If the processes that skilled readers use are established, then it will be easier to identify what beginning readers need to learn (Cain, 2010). To comprehend a

text, one needs to successfully read the word. Some words are already in the reader's lexicon, which is the place that the reader stores all of his/her known words, and some are not because the reader has never come across them before. So, the reader may or may not recognise the word (Cain, 2010). Skilled readers have the ability to read familiar words as well as words that they see for the first time. The Lexical Quality Hypothesis (Perfetti & Hart, 2002) provides an answer to what changes when there is increased exposure to printed words as it occurs with skilled readers. The answer is the lexical quality (Castles et al., 2018).

2.1.3 Lexical Quality Hypothesis

The *Lexical Quality Hypothesis* puts the weight of reading in the lexicon. The lexicon is the knowledge of the forms and meanings of written words which is also known as vocabulary knowledge (van den Bosch et al., 2019). Lexical quality is defined as how precise and flexible are the form and the meaning of a word's stored mental representation (Perfetti & Hart, 2002). Precision of the representation, which means knowledge of the exact spelling of the word, allows the reader not to confuse the word with other words that are spelled similarly and to directly access its meaning. Flexibility of the representation allows the reader to adapt to the different meanings a word may have (Castles et al., 2018). The basic idea of the Lexical Quality Hypothesis is that reading skill is supported by the reader's precision of the representation of orthography, phonology, morphology, meaning and the sheer number of know words (Verhoeven, Perfetti & Pugh, 2019). A word that is frequent in the reader's lexicon will be of high quality, but an unfamiliar word will be of low quality (Castles et al., 2018).

A high-quality word representation is one that integrates orthographic, phonological and semantic information. If any of these three types of information is deficient, the word representation will be of low quality and will result in poorer comprehension of the word (O'Connor et al., 2019). The more experienced the reader becomes with the print, the more the average of his/her lexical quality increases and this has as result that more cognitive resources are freed up and can be used for comprehension. When word decoding has become automatic and the words are recognized very rapidly, the reader's cognitive resources that are free can be used to gain comprehension. On the other hand, when decoding is not done automatically the cognitive resources are directed to this task, so less will be available for comprehension (Castles et al., 2018).

2.2 Reading comprehension

Reading comprehension is a prerequisite of everyday life, for professional success and for participation in society. The core of reading comprehension is to extract meaning from a text (Hjetland et al., 2017). The transition from learning to read to reading to learn is a crucial moment in the reader's development (Raudszus et al., 2019). According to the Rand Reading Study Group Report (2002 in Bernhardt, 2011 p.7) reading comprehension is defined as "the process of simultaneously extracting meaning through interaction and involvement with written language. We use the words extracting and constructing to emphasize both the importance and the insufficiency of the text as a determinant of reading comprehension". Another more detailed definition of reading comprehension comes from Johnston (1983 in Carrell, 1991 p.2-3) where he suggests that "reading comprehension is considered to be a complex behaviour which involves conscious and unconscious use of various strategies, including problem-solving strategies, to build a model of the meaning which the writer is assumed to have intended. The model is constructed using schematic knowledge structures and the various cue systems which the writer has given (e.g. words, syntax, macrostructures, social information) to generate hypotheses which are tested using various logical and pragmatic strategies. Most of this model must be inferred, since text can never be fully explicit and, in general, very little of it is explicit because even the appropriate intensional and extensional meanings of words must be inferred of their context".

The process of reading comprehension is described as an interactive process between bottom-up and top-down processing that requires both low-level and higher-level skills (Brevik et al., 2016). Low-level skills are those that facilitate word reading, for example phonological awareness and decoding and higher-level skills are those who support comprehension, such as vocabulary and comprehension strategies (Chung et al., 2019). Bottom-up processes involve word recognition along with grammatical information. Top-down processes refer to the creation of meaning with the interaction of text information, the reader's language knowledge and the reader's processing skills and strategies (Brevik et al., 2016). Different skills and knowledge have been found to correlate with reading comprehension, such as vocabulary, grammar, syntactic skills and cognitive ability (Cain, 2010; Chung et al., 2019; Hjetland et al., 2017).

2.2.1 Vocabulary

Vocabulary concerns the knowledge about the meaning of words (Cain, 2010). Vocabulary knowledge has been found to be essential for good reading comprehension since nobody can understand a text without knowing the meaning of the majority of words (Castles et al., 2018; Bernhardt, 2011; Chung et al., 2019; Brevik et al., 2016; Verhoeven, Voeten & Vermeer, 2019; Raudszus et al., 2018). Vocabulary is an unconstrained skill. Constrained skills are those that are learned quickly and eventually all of their elements are learned, for example letter knowledge. Unconstrained are the skills that take much longer developmental trajectory until they are mastered. Truth is that people never stop learning vocabulary, since there are always new words to learn (Cain, 2010). In fact, the relationship between vocabulary knowledge and reading comprehension is bidirectional. Vocabulary is a prerequisite for successful reading and reading comprehension, which, in turn, provides the opportunity to expand vocabulary knowledge (Castles et al., 2018).

Vocabulary knowledge is not just knowing the meaning of words. This aspect of vocabulary is called vocabulary breadth and it means the number of the words that a person knows. The other aspect of vocabulary is called vocabulary depth and it refers to knowing how individual words relate and associate with each other, how flexibly the reader can use them in any given text (Cain, 2010; Castles et al., 2018; Raudszus et al., 2018). The reader needs to have knowledge not only of single words but of multiword utterances, idioms and other figurative expressions (Castles et al., 2018). Research has shown that children's reading comprehension is more strongly associated to the depth than to the breadth of vocabulary knowledge (Cain, 2010). According to Cain (2010 p. 100) "growth in vocabulary involves adding words to a reader's lexicon, refining and consolidating the meanings of known words, and establishing and strengthening associations with words related in meaning".

Vocabulary can be measured in different ways. First, the reader could be asked to choose the correct meaning of a word, for example by pointing at the pictures that depicts this word. This is called receptive vocabulary. Alternatively, the reader could be asked to produce a word, for example use it in a sentence. This is called expressive vocabulary. Finally, the reader could be asked to define the meaning of a word (Cain, 2010).

However, the reader needs to know not only the meaning of individual words, but how they operate together in a sentence, to establish, in other words, the meaning of the sentence (Cain, 2010; Castles et al., 2018). The reader needs to establish the relation between word meaning

(vocabulary) and grammatical and syntactic structure in order to comprehend the meaning of the sentence. After the meaning of a sentence has been established, the reader needs to understand the whole text. For this to happen, some discourse skills are needed. Discourse are units of language that are longer than a sentence. To comprehend the text the reader needs to integrate the sentences of the text and then make inferences to fill in the gaps of what was not written (Cain, 2010). The understanding of the text needs to be monitored as well, which means to evaluate the comprehension of the text by using strategies such as looking back in the text to see if it makes sense or re-reading something that was not understood (Cain, 2010; Brevik et al., 2016). Finally, the reader needs to know how the ideas of the text are related to each other. These skills will help the reader build a coherent representation of the meaning of the text (Cain, 2010).

2.2.2 Cognitive resources

In addition to vocabulary knowledge, research has shown that cognitive resources also play an important role in reading comprehension (Hjetland et al., 2017). Some cognitive resources that have been found to influence reading comprehension are working memory, short-term memory, cognitive flexibility and inhibitory control (Castles et al., 2018). Working memory and short-term memory has been the most discussed among the reading comprehension research. Short-term memory is the mechanism that stores information over time when there are not other competing cognitive demands. Working memory is the mechanism that controls, regulates and maintains the information that are relevant to the task while engaging in other cognitively demanding tasks (Castles et al., 2018; Hjetland et al., 2017). How available is the working memory facilitates the building of richer, more detailed and well-connected representation of the text meaning. A greater working memory may retain more information and allow more inferences to be generated and more connections to be made. It could also deactivate irrelevant information, thus freeing resources that could be used in comprehension. Working memory has also been connected to vocabulary and inference making, which as seen earlier, are important factors of reading comprehension (Castles et al., 2018). Research as well suggests that other cognitive skills, such as nonverbal IQ, play a significant role in prediction of reading comprehension (Hjetland et al., 2017).

2.2.3 Language literacy

Language literacy in general has been found to contribute to reading and reading comprehension in both the first and the second language of a person. It is portrayed with different ways by different researches and there is not an agreement about how constructs associated with knowledge of the language should be represented (Lee & Scallert, 1997). Bernhardt (2011) points out that this is a catch-all term which involves all the knowledge about functions of written and spoken language and how sophisticated that knowledge is. It involves knowledge of vocabulary, grammar, pragmatics, metalinguistic knowledge and strategies. Another construct related to language literacy is language proficiency. It refers to language competence, metalinguistic awareness, and speaking, reading, listening and writing in appropriate ways depending on the context (Lee & Scallert, 1997).

2.2.4 The Reading Systems Framework

The *Reading Systems Framework* (Perfetti & Stafura, 2014) is a seminal model of reading comprehension which aims to capture its complexity (van den Bosch et al., 2019). According to the Reading Systems Framework (RSF) there are three constructs that underpin reading comprehension. The first one concerns the knowledge sources that are used, which are orthographic knowledge, linguistic knowledge and general knowledge which refers to general knowledge about the world. The second describes the processes that are involved in reading, which are decoding, word identification, meaning retrieval, sentence parsing, inferencing and comprehension monitoring. These knowledge sources can be used both in constrained ways and in interaction with each other. The third one captures the cognitive system that these processes take place which includes cognitive resources such as memory (Perfetti & Stafura, 2014; Castles et al., 2018; Li & Clariana, 2019; Raudszus et al., 2019; van den Bosch et al., 2019).

Central in RSF is the lexicon of the reader which mediates the interaction between the word identification system and the comprehension system (Li & Clariana). In RSF orthographic and phonological information are used to access the word's meaning. High-quality lexical representations allow for quick and accurate meaning retrieval. These representations contain information of orthographic, phonological and semantic characteristics. Eventually, the recognized words integrate into the context and lead to a representation of the sentence

which, in turn, enables text comprehension (Raudszus et al., 2018). In addition, high-quality lexical representation that allows for efficient word identification, enables freeing up of cognitive resources for integration processes of higher level (Raudszus et al., 2019). According to RSF lexical knowledge is a prime predictor of reading comprehension, in both children and adults, because of its critical role in word-to-text integration process (Li & Clariana). RSF can be used to identify weaknesses in the reading system that lead to failure in reading comprehension (van den Bosch et al., 2019).

2.2.5 Measuring reading comprehension

Assessment of reading comprehension is not an easy task. As Johnston (1983 in Carrell, 1991 p.3) points out "assessment of reading comprehension requires interpretation of an individual's performance of some task which is based on information from a given text within a given context. Thus, performance on the test will depend on characteristics of the text, the nature of the task, and the context, as well as the person's reading abilities and prior knowledge". Furthermore, the measures that are used to assess reading comprehension are not equivalent, in spite of tapping the same underlying construct. Different measures assess different types of information about reading comprehension and make different demands on skills of word recognition and language knowledge (Cain, 2010). There is a range of different respond formats that are used to measure reading comprehension including true/false or yes/no sentence judgements, multiple choice questions, cloze tasks and open-ended questions (Cain, 2010; Rayner et al., 2016).

When true/false sentence judgements are used the reader, after reading a text, is presented with a set of sentences and needs to judge whether the content of the sentences is true or false. The advantages of this type of measure is that no complex verbal response is needed, so there are low processing demands. In addition, it can be administered in big groups because readers can easily record their responses and it can be easily scored as well. However, this response format cannot guarantee that it will measure comprehension skills such as inference making, which needs to be spelled out in order to assess if one was generated or not (Cain, 2010).

With multiple choice respond format, the reader reads a text and then answers the question by selecting which of the offered alternatives is the correct one. These multiple-choice questions

probe the memory of the reader on what he/she just read (Rayner et al., 2016). This type of measure has higher processing demands than the true/false sentence judgement, since the reader needs to compare the different respond options. It is widely used to assess figurative language comprehension, for example idioms and expressions (Cain, 2010). Furthermore, this type of assessment is easy to score and can as well be administer in big groups (Rayner et al., 2016). However, the questions and the response options are very important and must be chosen with care, otherwise they could mislead the reader (Cain, 2010; Rayner et al., 2016). Finally, the multiple-choice format has the same limitation as the true/false sentence judgement in detecting generation of inferences (Cain, 2010).

Another type of measuring is cloze tasks. Cloze tasks are sentences with a missing word where the reader needs to select a word to replace it, usually from the given options. These sentences are either presented within a coherent text or on one sentence at a time. This response format can be as well administered in big groups of readers and scored easily. However, it demands very little attention because the reader could answer correctly by paying attention only to individual sentences and it does not reflect the ability to comprehend a whole text (Cain, 2010).

Another alternative measure format is open-ended question. After reading a text, the reader needs to answers a series of questions that test the understanding of the text and the reader's memory or write a summary of the text (Cain, 2010; Rayner et al., 2016). This type of measuring may provide a better estimation of reading comprehension compared to the others (Rayner et al., 2016). However, the reader needs to answer the questions by writing, which may underestimate the comprehension of someone with difficulties in expressive language and compromise the ability of younger children to write full responses because of their developing writing skills (Cain, 2010). Furthermore, scoring can be more subjective and scores may differ significantly depending on who is scoring it (Rayner et al., 2016). Finally, it is not easy to be administered in big groups because of the amount of time it needs to be answered and scored (Cain, 2010).

2.3 Second language reading and reading comprehension

Nowadays reading in more than one language is a prevalent reality for an increasing number of people all around the world. The reason of this increase is the need that arise from the

large number of people immigrating and the fact that travelling around the world and accessing unlimited sources of knowledge from every part of it is a lot easier than it used to be. English is now the most frequently learnt foreign language and it is characterised as the world's second language. Children in many countries learn English from a very young age and sometimes the acquisition of reading is simultaneous in English and the country's official language. There are also many multilingual countries where it is usual to acquire reading in multiple language in school (Saiegh-Hadad & Geva, 2010). The availability of internet the last two decades has also changed second-language reading a lot. It increased the number of people reading in a second language by making a huge amount of second-language material easy to access at any time and free of cost (Bernhardt, 2011). Because of this increased interest in second language reading, research has tried to provide a deep understanding of the processes of literacy acquisition in more than one language.

Reading was the only purpose of learning a foreign language in Europe and USA for centuries. Reading in a second language was a field of theoretical interest from the beginning of psychology and it became a research focus in the end of 19th and the beginning of the 20th century. However, reading in a second language was believed to be a derivative field that relied on beliefs, models and research designs of the first language (Bernhardt, 2011). Recent research has shown that reading in a second language is not an impoverished version of first language reading, but a process that requires unique reading abilities and lexical and grammatical flexibility (Bernhardt & Kamil, 1995).

Reading in a second language (L2) requires the involvement of two language systems. The reader needs to have access to both the first language (L1) and the second one. Reading in these two languages differs in many ways. In spite of the differences, it is important to understand the role that literacy in L1 plays in the reading development in L2. This role has been emphasized in research only recently (Jiang, 2011). Reading in L2 may refer to either becoming literate in school in another language than the one spoken at home, which is the case with immigrant children and it is characterised as emergent bilingualism, or reading in L2 when the reader has already acquired literacy in L1, which is the case with learning a foreign language. In the first case children need to learn to read in a language that they have yet to master, which could be a very challenging task (Verhoeven, Perfetti & Pugh, 2019). Whatever the context of learning, research has shown that there is a structural relation between reading comprehension in L1 and L2 (Brevik et al., 2016).

2.3.1 L2 vocabulary and L1 reading and vocabulary

L1 and L2 reading have the same set of components, but in L2 reading, L1 reading plays a significant role (Yamashita & Shiotsu, 2017). In both languages reading is comprised of many cognitive processes, such as decoding, vocabulary knowledge and syntactic processing (Jeon & Yamashita, 2014). It is well established that L2 vocabulary knowledge plays an important role in reading comprehension in L2 (Raudszus et al., 2018; Raudszus et al., 2019; Jeon & Yamashita, 2014; Yamashita & Shiotsu, 2017; O'Connor et al., 2019; Schaars et al., 2019; Verhoeven, Voeten & Vermeer, 2019). However, L2 vocabulary for L2 reading comprehension has received little attention by researchers until recently. This may be because of the complexity of lexicon. Every language contains thousands of words and for every word the reader needs to learn phonological, orthographic, morphological, syntactic, pragmatic, articulatory, idiomatic and semantic information (de Groot, 2006). Vocabulary knowledge is acquired continually. Its multidimensionality, such as receptive and productive vocabulary and vocabulary breadth and depth, is learned gradually each time the reader encounters a word (Jeon & Yamashita, 2014).

L2 vocabulary knowledge has been found to correlate more strongly with reading comprehension among adults than children (Jeon & Yamashita, 2014; Raudszus et al., 2018). An explanation of this might be the Matthew effect suggested by Stanovich (1986, 2000 in Jeon & Yamashita, 2014). He argued that vocabulary knowledge has a reciprocal causative relationship with reading comprehension. This means that readers with better vocabulary knowledge become better readers and through reading they acquire more vocabulary knowledge that facilitates further reading efficiency. On the other hand, poor readers will avoid reading and this will lead in poorer opportunity to gain more vocabulary knowledge.

When it comes to assessment of L2 vocabulary in research, vocabulary size is the preferred type of receptive vocabulary assessment. Productive vocabulary assessment is not frequently used. Receptive vocabulary tests often consist of tasks where a synonym or a definition that better matches the presented word should be selected. The words can be presented either within a text or a sentence or in isolation. The preference of receptive vocabulary assessment over productive vocabulary assessment may be due to the fact that reading itself is a receptive skill, so scores from a receptive test will correlate more strongly with scores from reading comprehension tests (Jeon & Yamashita, 2014).

However, although L2 vocabulary is a prerequisite for L2 reading comprehension, L1 vocabulary has also been found to have a positive effect on L2 reading comprehension (Schaars et al., 2019; Chung et al., 2019). In fact, it has been found to directly predict L2 reading comprehension (Raudszus et al., 2018). L1 vocabulary may help L2 acquisition by scaffolding. This means that already known concepts in L1, need nothing but a new label to be learned in L2. L1 vocabulary also indicates the conceptual richness and the amount of known words in L2 (Raudszus et al., 2018).

L1 reading has also been found to strongly correlate with L2 reading comprehension and be an important predictor of it (Brevik et al., 2016). L1 and L2 reading have many differences but in the same time they share some higher order processes and metalinguistic skills. These processes and skills are prone to be transferred from one language to the other (Jeon & Yamashita, 2014). This has inspired researches to investigate the process of this transfer.

2.4 Cross-linguistic transfer

Strong research evidence suggests that there is cross-linguistic transfer from L1 to L2 (Verhoeven, Perfetti & Pugh, 2019). There are different definitions of transfer. Haugen (1956 in Altmisdort, 2016 p.2) argues that transfer is "the overlapping of two languages". Odlin (1989 in Altmisdort, 2016 p.2) suggests that "transfer is the influence resulting from similarities and differences between the target language and any other language that has been previously acquired". When a person learns a language, he/she is interacting with the already acquired L1 skills which are transferring from L1 to L2 (Altmisdort, 2016). The nature of transfer is interactive. It is influenced by cognitive and linguistic factors, the distance between L1 and L2, language proficiency in both languages and language complexity of both languages. On the other hand, it is constrained by sociolinguistic and social-cultural factors. However, it is still unclear how the interaction of these factors takes place (Chung et al., 2019).

Transfer can be either positive or negative. Positive transfer occurs when previous knowledge of specific features acquired in one language contributes to easier learning of the same features in another language. Negative transfer occurs when knowledge of one language impedes acquisition of knowledge in the other language (Chung et al., 2019; Park, 2013). This happens when the two languages have very different structures (Chung et al., 2019). Furthermore, it is usually suggested that transfer is not reciprocal and it occurs from the

strong language to the weak one and not the other way around (Saiegh-Hadad & Geva, 2010). However, despite the fact that transfer from L2 to L1 is rarely studied, it has been found that there is some positive transfer of reading skills from L2 to L1 (Altmisdort, 2016).

Language distance between L1 and L2 is an important factor of cross-linguistic transfer (Chung et al., 2019). Transfer between L1 and L2 may differ depending on the linguistic distance of the two languages (Brevik et al., 2006). The hypothesis is that the shorter the distance, the more the transfer between these languages because of the similar simple processes such as decoding (Jeon & Yamashita, 2014). It has been observed transfer of morphological awareness and orthographic processing between languages that have similarities in these linguistic features (Chung et al., 2019). Transfer of cognate awareness has been observed in languages that are related etymologically (Chung et al., 2019; Melby-Lervåg & Lervåg, 2011). Cognates are words that in both languages derive from the same linguistic origin, have the same meaning and are orthographically and phonologically similar to each other (de Groot, 2006; Chung et al., 2019). Cognates can help with vocabulary acquisition in L2 due to more accurate L2 lexical processing (Jeon & Yamashita, 2014).

Language complexity may as well play an important role in cross-linguistic transfer (Chung et al., 2019). The orthographic distance between the two languages has been identified as key factor of transfer. The role of L1 reading in L2 reading can differ among languages depending on whether L1 is an alphabetic or a non-alphabetic language. The main focus of research has been on alphabetic L1 and L2 (Jiang, 2011). In this study two European, alphabetic languages are of interest. Some European languages have more shallow orthographies while others have deeper ones with more inconsistent grapheme-morpheme correspondences and morphological influences on spelling. In addition, some languages have simple and others complex syllabic structures (Seymour et al., 2003).

In the current study, readers have Norwegian as their L1 and English as L2. Both of them are Germanic languages and thus Norwegian may explain more variance in English reading comprehension than other alphabetic languages (Brevik et al., 2016). Norwegian have a relatively shallow orthographic depth and a relatively complex syllabic structure (Seymour et al., 2003). It is considered as a relatively consistent orthography (Lervåg & Aukrust, 2010). In contrast, English have a deep orthographic depth and a complex syllabic structure. It is considered as an inconsistent orthography (Seymour et al., 2003). The level of English proficiency among Norwegians has long been high. Norwegians are taught English as a

compulsory subject from the first until the eleventh year of the school. Then English are offered as an elective subject in the last two years of school. In fact, in a European reading assessment in English as L2 in 2000, Norway came in second among eight countries (Brevik et al., 2016).

2.4.1 Linguistic Interdependence Hypothesis

The Linguistic Interdependence Hypothesis (Cummins, 1979) is a framework for understanding cross-linguistic transfer that is influenced by cognitive psychology. It states that language skills in L2 is largely shared with linguistic ability in L1 (Cummins, 1979). L1 and L2 are interdependent and depends on common underlying proficiency (Melby- Lervåg & Lervåg, 2011). Proficiency acquired in L1 can transfer or facilitate learning in L2 (Chung et al., 2019). There are fundamental similarities in L1 and L2 skills which can be transferred from one to the other (Jiang, 2011). Specifically, L2 reading performance is largely shared with L1 reading ability. Once a language skill such as reading has been acquired in one language, it will not be reacquired in the second language (Bernhardt & Kamil,1995). Thus, we can hypothesize that skilled L1 readers will also read well in L2 (Yamashita, 2002). However, the Linguistic Interdependence Hypothesis does not take into account the importance of L2 language proficiency. It suggests that it is not as critical to the development of L2 reading as L1 reading abilities and that for example readers with low L2 proficiency can carry out academic reading tasks in L2 by using all of their L1 academic reading skills (Jiang, 2011).

2.4.2 Linguistic Threshold Hypothesis

The role of L2 language proficiency in L2 reading has been emphasized through the Linguistic Threshold Hypothesis (Cummins, 1979). The Linguistic Threshold Hypothesis assumes that the reader needs to develop a certain level of L2 language proficiency before L1 literacy skills transfers to L2 development (Cummins, 1979; Yamashita & Shiotsu, 2017; Bernhardt & Kamil, 1995; Yamashita, 2002; Brevik et al., 2006; Lee & Schallert, 1997; Jiang, 2011). In reading, a threshold has been connected with reader's language competence in L2 (Lee & Schallert, 1997). According to this hypothesis, even if a reader is skilled in L1 he/she cannot read well in L2 unless he/she reaches a certain level of L2 language proficiency (Brevik et al., 2006; Yamashita, 2002). This certain level is the threshold. Figure 2.2 shows

the concept of a threshold level in L2 reading in correlation with L1 and L2 reading performance at different levels of L2 proficiency. According to the figure, below a certain level of proficiency, in this case level 3, there is no relation between L1 and L2 reading, whereas beyond that level, the correlations between reading in the two languages increases (Lee & Schallert, 1997).

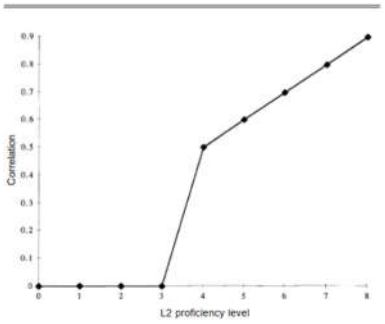


Figure 2.2 Hypothetical model of the Threshold Hypothesis (Lee & Schallert, 1997 p.3)

Clarke (1978), in accordance with the Linguistic Threshold Hypothesis, argued as well that there is a certain point of L2 language proficiency, which he named linguistic ceiling, below of which transfer of L1 reading skills is not possible. Good L1 reading skills cannot compensate when reading in L2 because the lack of L2 language proficiency will "short-circuit" the L1 reading ability. Within this threshold or linguistic ceiling existence hypothesis is the notion that language is a key factor of reading ability, which means that if on does not know the language, reading in this language is impossible (Bernhardt & Kamil, 1995). So, the argument is not whether there is transfer between L1 and L2 but when this transfer occurs (Jiang, 2011).

Research suggest that L2 proficiency tends to be a better predictor of L2 reading than L1 reading ability for learners who are not yet skilled. When the reader becomes more proficient in L2, L1 becomes a lot more important and this leads to successful transfer of reading skills from L1 to L2. However, this threshold level of L2 proficiency is not constant. The relationship between L1 reading, L2 proficiency and L2 reading is dynamic and interactive and depends on different factors, such as the reader's level of reading ability and the type of the reading task (Jiang, 2011).

2.4.3 A reading or a language problem?

These two transfer hypotheses have been well summarised by a research question. Alderson (1984 in Yamashita & Shiotsu, 2017) addresses the questions of whether L2 reading is a reading problem or a language problem. Basically, what he was asking was whether the individual differences in L2 reading comprehension are explained better by the individual differences in L2 knowledge, such as vocabulary or by differences in more cognitive and meta-cognitive processes, such as working memory (Jeon & Yamashita, 2014). In other words, whether L2 reading difficulties are better explained by limited L2 knowledge or by weak L1 reading ability (Yamashita & Shiotsu, 2017). He concluded that "it appears to be both a language problem and a reading problem, but with firmer evidence that it is a language problem, for low levels of L2 competence, than a reading problem" (Alderson, 1984 in Lee & Schallert, 1997 p.2). This is in line with Jeon's and Yamashita's (2014) finding that it is seems to be a language problem, although it is still significantly a reading problem due to the strong correlation of L2 reading comprehension and L1 cognitive and meta-cognitive processes.

2.4.4 Compensatory theory

The transfer of skills from L1 to L2 and the role of L2 language proficiency in L2 reading is also discussed in Bernhardt's (2011) compensatory theory. She proposed a model of L2 reading where the component that contributes the most is L2 language knowledge, by which she refers to vocabulary knowledge, grammatical forms, cognates, L1-L2 distance etc. L2 language knowledge accounts for up to 30% of L2 reading comprehension. The second largest contribution was that of L1 literacy. This refers to vocabulary, alphabetics, text structure etc in L1 and it accounts for up to 20% of L2 reading comprehension. The remaining 50% is unexplained variance which she argues must surely entails factors such as motivation and background knowledge. She proposed that compensation does not mean that these components are independent of one another, rather that they are "more than dependent, they are inextricably intertwined because they are used by readers simultaneously in a compensatory fashion" (Bernhardt, 2011 p. 63). She argues that all these information sources are constantly interacting with each other, but none of them is dominating, rather they are supporting each other (Yamashita & Shiotsu, 2017). Thus, according to the compensatory model of L2 reading, if the reader has a weakness in one area, it will be compensated for by knowledge from another area (Brevik et al., 2016). Schaars et al. (2019) also suggests that

when L2 readers face a lack of semantically related skills, they compensate by relying on phonologically, orthographically or memory related skills that are better developed. However, more research is needed to support this suggestion.

Cross-linguistic transfer is a complex process determined by multiple factors. It is still unknown how these factors interact with each other. In spite the evolution of many transfer theories, none of them can "provide a consistent and comprehensive account of the empirical evidence available" (Chung et al., 2019 p.10). In addition, transfer of constructs such as reading comprehension strategies and orthographic processing has yet to be understood.

3. Method

3.1 Design

The goal of the present study is to examine whether L2 (English) vocabulary accounts for L2 reading comprehension, after accounting for L1 reading comprehension, in native Norwegian speakers. The study is based on data collected for the Norwegian part of an international study about eye-movements of adults during reading in many languages with different writing systems. The Center for Advanced Research in Experimental and Applied Linguistics at McMaster University in Canada is responsible for the international study named Multilingual Eye-Tracking Corpus (MECO) (Cop et al., 2017). To examine the question of the study a quantitative method was used and the participants underwent various psychometric tests. No variables were manipulated during the study which is based on natural variation among participants and the study is non-experimental (Tabachnick & Fidell, 2013).

3.2 Participants

There were 57 participants that took part in the study. Due to unmet criteria, specifically having leaved in an English-speaking country for more than 6 months or having vision or reading difficulties, 6 of them and because of missing data 1 were excluded. Thus, data from 50 participants were used in the present study. Participants should be from 17 to 30 years old. They should have Norwegian (Bokmål) as their mother tongue and be proficient non-native English speakers. However, they should not have lived in an English-speaking country for more than 6 months. Other participation criteria included not using glasses and not having diagnosis of vision, auditory or reading disability. Selection of the participants was made informally among the acquaintances of the master students that are working with the project. The 50 participants have different socio-economic and educational background.

3.3 Measures

The measures that were used in this project were standardised tests or parts of them and tests that were developed for the purposes of the study. The tests were measuring knowledge of vocabulary in Norwegian and English and reading comprehension in both languages. Testing took place from 01.12.2019 until 31.01.2020. Participants read passages while their eye-movements were tracked by an eye-tracking device and were administered various individual

differences tests both in L1 and L2 as part of the project. The whole testing process took about 1.5 to 2 hours for each participant depending on their skills. The individual differences tests included a non-verbal IQ test, a motivational scale and tests in sight word efficiency and phonemic decoding efficiency, vocabulary and spelling in both languages. Testing was completed by 3 master students in the Department of Special Needs Education in the University of Oslo.

3.4 Description of the variables

Tests that measure knowledge of vocabulary in Norwegian and English and reading comprehension in Norwegian have been used as independent variables and a test that measures reading comprehension in English as the dependent variable. Vocabulary in Norwegian was measured with the Norwegian version of the British Picture Vocabulary Scale (BPVS) (Dunn, Dunn, Whetton, & Burley, 1997) and vocabulary in English with a Vocabulary size test adopted from Nation and Beglar (2007) and with LexTale in English (Lemhöfer & Broersma, 2012). To measure reading comprehension in Norwegian a test with comprehension questions about texts that had just been read were developed especially for this project. For English reading comprehension measuring texts and questions were selected from sample materials for larger comprehension test, the Reading Comprehension segment of the ACCUPLACER test for colleges.

3.4.1 Vocabulary in English

Two tests were used to measure vocabulary in English. The first one is an online Vocabulary size adopted from Nation and Beglar (2007). It is a receptive vocabulary test with 14.000 words in the original version which contains 140 multiple choice items with 10 items from each 1000-word family level. Each item contains one of these words together with an example of how the word is used in context. Four definitions are given for each item and the participant should select the correct one without a time restriction. This original test was adapted to enable more rapid assessment. The first 10 items that represented the first 1000-word family with the most common English words were skipped and the participant started from the 2000-word family. If the participant answered correctly 5 or more of the 10 items in each level, he/she was allowed to continue to the next word family level. The test stopped if

the participant had more than 5 wrong answers in any given word family level. The score is the number of correct responses.

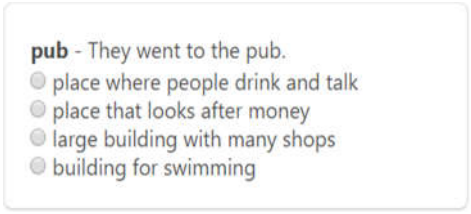


Figure 3.2 Example of an item from the 2000-word family

The second one is LexTale (Lemhöfer & Broersma, 2012) which is a lexical decision test. It consists of 60 items that were presented on the screen. The participant had to decide if each of them was a real existing English word or not (40 were real words and 20 were non-words). If they thought that the word existed, they had to click on the "yes" button and if not on the "no" button. British English spelling was used in this test which was an online test and was not timed as well. The score is the percentage of correct responses, corrected for the unequal proportion of words and non-words in the test by averaging the percentages correct for these two item types.

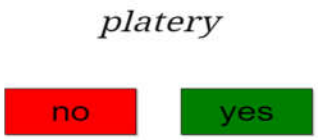


Figure 3.3 Example of a word from LexTale

3.4.2 Reading Comprehension

To measure reading comprehension in both languages two tests were developed for this project. The tests used for English were the same for all countries taking part in the international study. The tests used for the first language were different for each country but created in the same way. Each test consists of twelve texts and some comprehension questions to each of them. The participant read each text while the eye-movements were tracked. After each text the participant answered the questions that follow. There was no time restriction in reading the texts or answering the questions.

In the L1 test participants read twelve texts in Norwegian. Participants had to read each text silently and then answer four yes/no questions (Figure 3.4). These texts were given to all

participating countries in English. They were taken from Wikipedia entries and were edited for length and complexity. Five of them (texts number 1,3,7,11 and 12) should be translated in the official language of each country and the remaining should be adapted in this language. The five texts were translated into Norwegian with the same number of sentences and the same general content (Appendix 1). The translation was done by native Norwegian speakers that did not participate in the study and they were checked and approved by other native Norwegian speakers as "naturally read in Norwegian". More than one person checked and approved each text. The reading comprehension questions were as well translated in Norwegian from the English prototype.

The remaining 7 texts were adapted in Norwegian by the three master students that are working with the project. The adaptation was made using Wikipedia articles or similar material in Norwegian keeping the same or similar topic that would be familiar to the Norwegian culture and the same complexity (Appendix 1). The material that was used was either under the Creative Commons license or permission was asked and given by the authors. The reading comprehension questions were made by the master students, since the texts were adapted and the original questions could not be preserved. The adapted texts were as well read and approved as "naturally read in Norwegian" by native Norwegian speakers who did not participate in the study. All native Norwegian speakers that were used were acquaintances of the master students. The score is the percentage of correct answers in all twelve texts' questions.

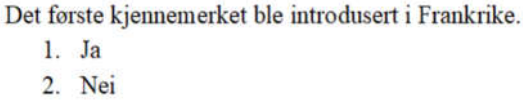


Figure 3.4 Example of reading comprehension questions in Norwegian (text n.12)

In the L2 test participants read twelve texts in English after they completed the Norwegian part of the study. These texts represented a range of reading complexity from grade 8 of an English-speaking school to college level (Appendix 2) and they varied in length as well (from 6 lines in text number 2 to 12 lines in text number 5). After silently reading each text while the eye-movements were tracked, the participant had to answer two multiple choice comprehension questions (Figure 3.5). The topics of the texts were various, for example there were texts about Samuel Morse and Leonardo da Vinci, the different types of intelligence, sleep's benefits, leeches and technological progress. Texts and questions were selected from

sample materials of the Reading Comprehension segment of the ACCUPLACER test for colleges. All texts for L2 were the same for all participating countries and they could not be altered or adapted. The score is the percentage of correct answers in all questions.

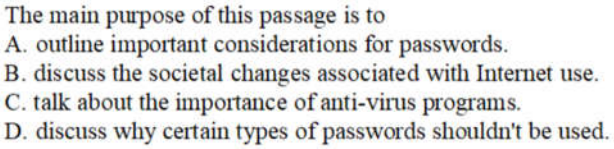


Figure 3.5 Example of reading comprehension questions in English (text n.5)

3.5 Analysis

To analyse the collected data the statistical program Jamovi 1.1.9.0 was used. First a descriptive analysis was done to provide an insight in the characteristics and distribution of every variable. After that a correlation analysis was done to determine the correlation between the variables. Finally, multiple regression was used to assess the relationship between the dependent and the independent variables. This type of analysis allows to assess whether knowledge of English vocabulary and knowledge of Norwegian vocabulary explain the reading comprehension in English for adults that are fluent in English.

3.6 Validity and Reliability

To judge the quality of every research project it is crucial to follow the principles and standards that measure whether it is of sufficient quality or not (Cozby, 2015). There are other criteria as well, but validity and reliability are among the most important ones (Gall et al., 2007).

Validity refers to the degree that the variables measure what they are purported to measure (Cozby, 2015). Validity is a function of the empirical method procedures used in the study (Lund, 2002). According to Cook and Campbell (1979) there are four types of validity that refer to different aspects of the research. These types of validity are Statistical Conclusion Validity, Internal Validity, Construct Validity and External Validity.

Statistical Conclusion Validity refers to whether it is reasonable to presume that the relationship between the independent and dependent variables is statistically significant, so if there is covariation (Cook & Campbell, 1979). There are two types of error that can occur

regarding statistical significance of the variables' relationship. Type I error refers to falsely rejecting the null hypothesis. When Type I error occurs, the conclusion is that there exists covariation between the variables when it does not. Type II error refers to falsely not rejecting the null hypothesis. When Type II error occurs, the conclusion is that there exists no covariation when it does. The likelihood of Type II error increases when the sample size is small. Type I and Type II errors are major threats to this type of validity. (Cook & Campbell, 1979). Sampling error is another cause of false result in statistical significance along with small sample size which can affect the statistical conclusion validity of the research (Cook & Campbell, 1979).

Internal validity refers to the evidence that the research is a true and accurate reflection of the participants, procedures and the setting observed (Dollaghan, 2007). Cook and Campbell (1979) refer to internal validity as the one measuring whether the variables are causally related to one another in the way that they were manipulated or measured. After establishing covariation between two variables there is need to decide if there is a causal relationship between them and the direction of this causality. Causality between the variables can be assessed when there is a real correlation of them. This means that it needs to be ascertained that their correlation is not related to a third variable (Cook & Campbell, 1979). Direction of causality is usually assessed through time sequence. Time sequence is often available for experiments but not for observational studies because in this type of studies the statistical correlation can have many possible causalities and time sequence may cannot be observed (Cook & Campbell, 1979).

According to Cook and Campbell (1979) construct validity refers to whether we can make generalisations of higher-order constructs from research operations. Construct validity refers to whether the operationalizations on the variables measure the relevant constructs (Lund, 2002). Construct validity is also about what is called confounding. Confounding occurs when it is not possible to disentangle the effects of two or more processes (Dollaghan, 2007). In research there are some theoretical constructs the effects of which are being studied. Confounding in research means that a relationship between two or more theoretical constructs can be interpreted in different ways by different researchers. Thus, construct validity refers to these central constructs of the study and the way their relationship is interpreted. Many studies have abstract constructs the relationship of which can be interpreted in various ways. Therefore, these constructs should be operationalised in a way that they reflect what is actually being studied. Construct validity refers to the compliance between the

operationalised constructs of the study and the theoretical content of the constructs, so whether the operationalised constructs refer to what is actually being studied (Cook & Campbell, 1979).

External validity refers to the degree to which the results of a research can be generalised to the population. Whether the results are a true and accurate reflection of participants, procedures and settings other than those observed or not is what external validity refers to (Dollaghan, 2007). The goal of every research is to have results that can be generalised in other people, settings and times. For this to happen the sample that was used to conduct the study should be representative of the whole population. Sampling method, sample size and homogeneity in sample are crucial for a good external validity. External validity can be enhanced more by a group of smaller studies with random samples than by only one big study with representative sample (Cook & Campbell, 1979).

Reliability is the degree to which an assessment tool produces stable and consistent results. In qualitative research reliability refers to the degree to which research is accurate and credible (Cozby, 2015). According to the COSMIN Taxonomy of Measurement Properties (COSMIN, 2018) reliability refers to the degree to which the measurement is free from measurement error, and it contains the measurement properties internal consistency and measurement error.

Internal consistency is the degree of interrelatedness among the items of a test. If all items on a test measure the same construct or idea, then the test has internal consistency reliability. A low internal consistency means that there are items or sets of items which are not correlating well with each other. Cronbach's alpha is a measure of internal consistency, that is, how closely related a set of items are as a group (COSMIN, 2018). There are different opinions on the value of sufficient reliability coefficient. Gall et al. (2007) propose .80 or higher as sufficiently reliable for most research purposes. Measurement error is the difference between a measured quantity and its true value. It includes random error which is a naturally occurring error that is to be expected with any experiment and systematic error which is caused by a mis-calibrated instrument that affects all measurements (COSMIN, 2018).

3.7 Research ethics

In order for any research to promote good scientific practice and to avoid scientific misconduct, research ethics are applied. Research ethics are norms, values and arrangements

that regulate scientific activities and are based on general morality in science and society (The Norwegian National Research Ethics Committees [NESH], 2016).

The present project has been approved by the Norwegian Centre for Research Data (NSD) which is responsible for the use of personal information, the data that are collected and the research ethics. Before the beginning of the testing a consent form (Appendix 4) with all the necessary information was given to each participant. There was stated what the purpose and the procedure of this study is, the rights the participant has, who is responsible for this study, the way that the participant's information will be stored and used. Afterwards the participant gave consent stating that he/she agrees on taking part in the study and that the data will be used anonymized outside EU where they will be available for future research. The consent was given by checking boxes for every part that the consent was needed and not by signing the consent form.

Anonymity is guaranteed in the project. No information about the name, the date of birth or the address were collected. That is the reason why participants could not sign the consent form. A demographic questionnaire was administered which was adapted from the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al., 2007). The data collected from this questionnaire were about age, gender, years of education and language history. A unique code was used as an ID for each participant in order for the data to be unidentifiable and anonymized. The data collected will be shared publicly and will be stored for future research. A positive environment was created during testing for all participants. They were encouraged to take a break in the middle of the session or more whenever they felt tired. They were also encouraged to make themselves comfortable by adjusting the chair and the table they were using during eye-tracking depending on their height and were offered eye-drops in case their eyes were tired.

4. Results

In this chapter there is going to be analysis of the study's results. First the descriptives of the variables will be presented and then their deviation from normal distribution will be mentioned. Next there will be presentation of the reliability of the tests that were used. Finally, the results from the correlation analysis and the regression analysis will be presented.

4.1 Descriptive analysis of the variables

We use descriptive statistics to describe samples of the populations in terms of variables or combination of variables and they provide estimations about the central tendency in the population (Tabachnick & Fidell, 2013). The mean is the average of the scores and is strongly influenced by very low and very high points. The median on the other hand, is the value below and above which half of the scores lie and it requires only ordinal scaling so it is not influenced from very low or very high points. The standard deviation shows how spread the data are and is very strongly influenced by very low or very high points as well (Cumming & Calin-Jageman, 2017).

Screening for normality of variables is important because the analysis is often better if all variables are normally distributed. Normality can be assessed by statistical or graphical methods through skewness and kurtosis which are two of its components (Tabachnick & Fidell, 2013). Skewness is about the symmetry of the distribution. A variable is skewed when its mean is not in the centre of the distribution. The distribution can be either positively or negatively skewed (Cumming & Calin-Jageman, 2017). Positively or to the right skewed is when the frequent scores are clustered at the lower end while the tail points to the more positive scores. Negatively or to the left skewed is when the frequent scores are clustered at the higher end while the tail points to the lower scores (Field, 2016). Kurtosis is about peakedness of the distribution, whether there are too many or too few scores in the ends of the distribution (Tabachnick & Fidell, 2013). A distribution has positive kurtosis when there are too many scores in the tails, is called leptokurtic and it tends to look pointy, whereas it has negative kurtosis when there are too few scores in the tails, is called platykurtic and it looks a little bit flat (Field, 2016). In normal distribution the values of skewness and kurtosis are zero. The farther from the zero skewness and kurtosis of a distribution are, the bigger the deviation it has from the normal distribution (Tabachnick & Fidell, 2013).

Table 4.1 shows an overview of the variables that were used in the analysis. It shows their mean, median, standard deviation, skewness and kurtosis. A more detailed description of the variables will follow.

	Mean	Median	SD	Skewness	Kurtosis
Vocabulary size	51.80	48.50	20.70	-0.121	-1.040
LexTale	75.60	76.30	11.40	0.085	-0.660
Reading comprehension L1	0.72	0.71	0.09	-0.023	-0.676
Reading comprehension L2	0.66	0.69	0.15	-0.609	-0.457

Table 4.1 Descriptives of the variables

4.1.1 Assessment of the variables that represent vocabulary knowledge

Vocabulary knowledge is represented by two variables in this study. Vocabulary size and LexTale, represent vocabulary knowledge in English.

Assessment of the variable Vocabulary size

Vocabulary size tests participant's receptive vocabulary. It consists of 13 groups of 10 items each, so 130 items in total. Each item has 4 multiple choice answers so the participant had 25% chance to randomly choose the correct answer which is equivalent to 32.5 correct answers. The mean score of the test was 51.8 which is equivalent to 39.8% correct answers. 14% of the participants had a score equivalent or less than 25%. This means that the scores are not a result of guessing. There is no ceiling or floor effect. Skewness is −0.121 and the distribution is negatively skewed. Kurtosis is −1.040 so the distribution is platykurtic and has a moderate deviation from normal distribution.

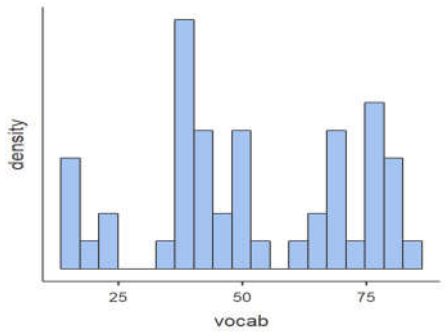


Figure 4.1 Vocabulary size

Assessment of the variable LexTale

LexTale is a lexical decision scale. It consists of 60 items with 2 possible answers each. Thus, the participant had 50% chance to randomly choose the correct answers which is equivalent to 30 correct answers. The mean score of the test is 75.6% correct answers which is equivalent to 45.4 correct answers. There was no participant who answered 50% or less of the items correctly and that indicates that the scores were not a result of guessing. Skewness is 0.085 which means that the distribution is very slightly skewed to the right. Kurtosis is −0.660 so the distribution has a small deviation from the normal distribution.

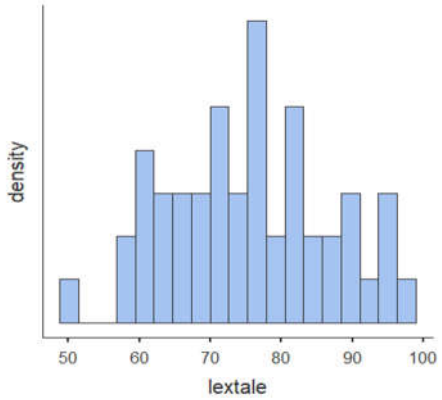


Figure 4.2 LexTale

4.1.2 Assessment of the variables that represent reading comprehension

Reading comprehension is represented by two variables. One is reading comprehension in L1 which is Norwegian and the other is reading comprehension in L2 which is English.

Assessment of the variable reading comprehension in L1

This test consists of 12 texts with 4 yes or no questions, thus 48 questions. The participant had 50% chance to randomly choose the correct answer which is equivalent to 24 correct answers. The mean score is 0.72 - 72% - which is equivalent to 34.5 correct answers. There was only one participant that had a score equal to 50% or lower. The value of skewness is −0.023 so the distribution is very slightly skewed to the left. Kurtosis is −0.676. The distribution has a small deviation from normal distribution. There is no ceiling or floor effect.

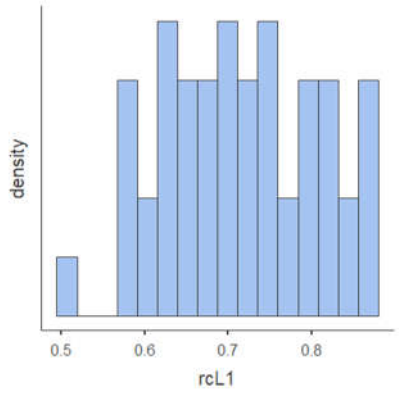


Figure 4.3 Reading comprehension in L1

Assessment of the variable reading comprehension in L2

This test also has 12 texts but instead of 4 yes/no question, each text has 2 questions with 4 possible answers each. Thus, there are 24 questions and the participant had 25% chance to choose the correct answer randomly which is equivalent to 6 correct answers. The mean score of the test is 0.66 -66% - which is equivalent to 15.9 correct answers. There was no participant with a score equal to 25% or less. The distribution is skewed to the left with −0.609 skewness. There is a ceiling effect since 32% of the participants had 18 or more correct answers and 14% of the participants had 19.2 or more correct answers. Kurtosis is −0.457, thus the distribution has a small deviation from normal distribution.

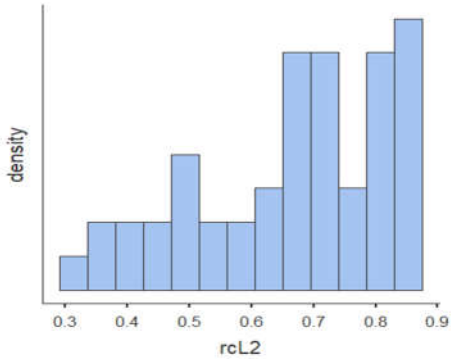


Figure 4.4 Reading comprehension in L2

The distributions of all variables had a skewness and kurtosis that was below or above zero. To test normality of distributions the Shapiro-Wilk test will be used. The null hypothesis in this test is that the population is normally distributed. If the p value is less than the chosen alpha value, which usually is 0.05, the null hypothesis is rejected and the data are not normally distributed. If the p value is greater than the alpha value, the null hypothesis cannot be rejected and the data are normally distributed (Field, 2009). As seen in Table 4.2 only the variables Vocabulary size and Reading comprehension in L2 have a p value lower than 0.05. Concluding, the variables have either a non-significant or a small deviation from normal distribution.

	Shapiro-Wilk p-value
Vocabulary size	0.005
LexTale	0.725
Reading comprehension L1	0.329
Reading comprehension L2	0.015

Table 4.2 Results from Shapiro-Wilk test

4.2 Reliability of the variables

Reliability is the degree to which an assessment tool produces stable and consistent results (Cozby, 2015). In this study reliability is measured with Cronbach's alpha (α). Cronbach's α is a function of the number of items in a test, the variance of the total score and the average covariance between item-pairs (Cronbach, 1951). The theoretical value of Cronbach's α varies between 0 and 1. Higher values of α are more desirable. There are different opinions on the value of sufficient reliability coefficient but a reliability of .70 or higher is usually reasonable. The following table shows the α value for each variable.

	Cronbach's α
Vocabulary size	0.98
LexTale	-
Reading comprehension in L1	0.62
Reading comprehension in L2	0.68

Table 4.3 Reliability of variables

As is seen from the table, tests that measure knowledge of vocabulary have a high reliability coefficient. Test Vocabulary size that measures vocabulary in English have α=0.98. The reliability coefficient for LexTale was taken from Lemhöfer and Broersma's (2012) study. In this study, which consisted of five parts assessing differing aspects of English skills, 72 Dutch and 87 Korean native speakers participated. The first part was LexTale, the same material that was used for the present study as well. The scoring was made with three different ways and reliability coefficients were reported for all three of them. In the present study just the reliability coefficient that corresponds to the scoring method that was used is reported. Furthermore, Lemhöfer and Broersma report separate reliability coefficients for the group of Dutch and the group of Koreans. In the present study the one from the group of Dutch is reported because of the similarity of the Norwegian language to the Dutch in comparison to the Korean. The reliability coefficient that was used was not Cronbach's α, but split-half reliability which was 0.81. Reliability coefficients in tests that measure reading

comprehension are lower though. Reading comprehension in L2 has α=0.68 which approximates closely the sufficient value of 0.70, but is a moderately sufficient internal consistency. Reliability coefficient in L1, α=0.62, similarly to the one in L2 approximates closely the sufficient value, however it indicates moderately sufficient internal consistency as well.

4.3 Bivariate Correlation analysis

Correlation is used to measure the relationship between the variables. It is a measure of the size and direction of the relationship between two variables and squared correlation is a measure of how strong is the association between them (Tabachnick & Fidell, 2013). A bivariate correlation is a correlation between two variables (Field, 2009). Pearson's r is the most frequently used correlation coefficient. It can range between the values of −1 and +1. Values close to .00 represent no linear relationship or predictability between the variables. Values of +1 or −1 indicate perfect predictability between variables (Tabachnick & Fidell, 2013). A coefficient of −1 indicates a perfect negative relationship which means that if one variable increases, the other decreases by a proportionate amount. A coefficient of +1 indicates that the two variables are perfectly positively correlated, so as the one increases, the other increases as well by a proportionate amount (Field, 2016). Values of +0.1 or −0.1 represent a small effect, +0.3 or −0.3 a medium effect and +0.5 or −0.5 a large effect (Field, 2016). Since some of the variables used in the present study are not normally distributed, Spearman's rho (ρ) is going to be used also as correlation coefficient. Spearman's ρ is a non-parametric statistic that can be used when the data have violated parametric assumptions, in this case normality (Field, 2009). Both Pearson's r and Spearman's ρ will be reported (Table 4.4).

The correlation analysis showed that reading comprehension in L2 has a strong correlation with almost all independent variables. The correlation coefficient was higher than +0.5 with Vocabulary size which measures vocabulary in English (ρ=0.51) and reading comprehension in L1 (ρ=0.510). There is a medium correlation between reading comprehension in L2 and LexTale which is the second measure of vocabulary knowledge in English (ρ=0.44), which approximates strong correlation though. When the independent variables are very strongly related, r > 0.90, then the problem of multicollinearity can occur. When multicollinearity occurs, the shared variance of the variables becomes larger and the contribution of each

variance in explaining the relationship is not unique which leads to unclear explanation of the relationship (Tabachnick & Fidell, 2013). There is no multicollinearity between the variables in this study. Bivariate correlation analysis shows whether there is a relationship between two variables and the degree of this relationship, but it does not show if the one variable can be predicted from the other (Tabachnick & Fidell, 2013). In order to show if vocabulary knowledge in English can predict reading comprehension in English, after controlling for reading comprehension in Norwegian, for native Norwegian speakers, a regression analysis needs to be done.

Correlation Matrix

		Vocabulary size	**Reading comprehension in L1**	**Reading comprehension in L2**	**Lextale**
Vocabulary size	r / ρ	-	**0.367** **	**0.51** ***	**0.617** ***
	p-value	-	**0.009**	**< .001**	**< .001**
Reading comprehesnion in L1	r / ρ	0.292 *	-	**0.564** ***	**0.376** **
	p-value	0.04	-	**< .001**	**0.007**
Reading comprehension in L2	r / ρ	0.545 ***	0.52 ***	-	**0.371** **
	p-value	< .001	< .001	-	**0.008**
LexTale	r / ρ	0.617 ***	0.359 *	0.44 **	-
	p-value	< .001	0.011	0.001	-

Note. * p < .05, ** p < .01, *** p < .001

Table 4.4 Bivariate Correlation Matrix. Spearman's ρ is noted in bold text above the diagonal

4.4 Hierarchical multiple regression

Regression analysis allows to assess whether values of a dependent variable (DV) can be predicted from one or more independent variables (IVs) (Tabachnick & Fidell, 2013). Predicting a dependent variable, also called outcome variable, from several independent variables, also called predictor variables, is called multiple regression (Field, 2009). In hierarchical or sequential regression, the order that the predictor variables (IVs) are entered in the equation is specified by the researcher according to logical or theoretical considerations (Tabachnick & Fidell, 2013). Already known predictors from previous research should enter the equation first and then the researcher can add any new predictors (Field, 2009). Hierarchical regression allows to test the proportion of variance attributed to some predictors after accounting for variance due to predictors already in the equation (Tabachnick & Fidell, 2013). A hierarchical multiple regression analysis has been conducted with reading comprehension in L2 as the outcome variable (DV) and the remaining as the predictor variables (IVs).

4.4.1 Assumption checking

To generalise the results of a regression analysis done on a sample to a population several assumptions should be true (Field, 2009). Examining the residual plots provides a test of assumptions between outcome and predictor variables (Tabachnick & Fidell, 2013). Residuals are the differences between the predicted values of the outcome variable and the observed values of the outcome variable. If a model fits well the sample data then all residuals will be small, but if it does not then the residuals will be large (Field, 2009). Residual plots (Appendix 3) should be tested for assumptions of normality, linearity and homoscedasticity (Tabachnick & Fidell, 2013).

The assumption of normality is that the errors of prediction are normally distributed around each predicted outcome score (Tabachnick & Fidell, 2013). This means that the residuals are often very close to zero and that much greater differences happen occasionally (Field, 2009). There should be a pileup of residuals in the centre of the residual plot at each value of predicted score and the normally distributed residuals should be trailing off symmetrically from the centre (Tabachnick & Fidell, 2013).

The assumption of linearity is about the linear relationship between predicted outcome scores and errors of prediction. If there is nonlinearity, the shape of the scatterplot will be curved

instead of rectangular (Tabachnick & Fidell, 2013). If a non-linear relationship is modelled with a linear model then generalizability of the findings will be limited (Field, 2009). However, nonlinearity does not mean that the analysis is completely invalid, rather it is weakened. The analysis cannot map the full extent of the relationships between the outcome and the predictor variables (Tabachnick & Fidell, 2013).

The assumption of homoscedasticity is that the residuals at each level of the predictor variables should have the same variance (Field, 2009). It means that the band of the residuals is approximately equal in width at all values of the outcome variable (Tabachnick & Fidell, 2013). When the variances of the residuals are very unequal there is heteroscedasticity (Field, 2009). Heteroscedasticity occurs when the band of the residuals becomes wider at larger predicted values of the outcome variable (Tabachnick & Fidell, 2013).

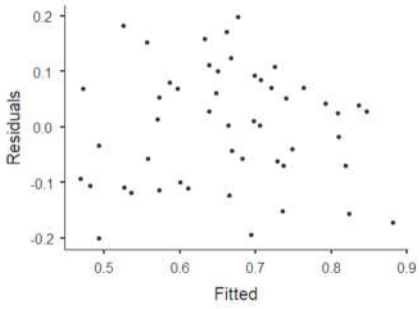


Figure 4.5 Model's residual plot

No perfect multicollinearity is an assumption as well. It is assumed that there is no perfect linear relationship between two or more predictor variables, they should not have a very high correlation (Field, 2009). In regression analysis multicollinearity can also be signalled by very large standard errors for regression coefficients because when the Pearson's r is 0.9 the standard errors of the regression coefficients are doubled. If multicollinearity occurs, none of the regression coefficients may be significant because of the large size of the standard errors (Tabachnick & Fidell, 2013). The variance inflation factor (VIF) indicates whether a predictor variable has a strong relationship with the other predictor variables. It is suggested that a value of 10 of VIF should raise concern about multicollinearity. The tolerance statistic is another indicator of linearity. Tolerance statistic is reciprocal to VIF (1/VIF) and values below 0.2 should cause concern (Field, 2009).

Collinearity Statistics		
	VIF	**Tolerance**
Reading comprehension in L1	1.29	0.772
Vocabulary size	2.29	0.437
LexTale	1.78	0.561

Table 4.5 Variance inflation factor (VIF) and Tolerance

When these assumptions are met, the coefficients and parameters of the regression are said to be unbiased. An unbiased model shows that on average the sample regression model is the same as the population model. This means that even if the assumptions are met, it is possible that a model from a sample may not be the same as the population model, but there is an increased possibility that it is (Field, 2009).

4.4.2 Regression analysis

Regression analysis was done to study in what degree does English vocabulary predict reading comprehension in English, after accounting for reading comprehension in Norwegian, for Norwegian native speakers. Therefore, reading comprehension in L2 was the outcome (dependent) variable. The other variables were added in two blocks. Reading comprehension in L1 was added in the first block of the model. Measures of vocabulary knowledge in English, Vocabulary size and LexTale, were added in the second block.

The assumption of normally distributed residuals was checked. Figure 4.6 shows the standardised residuals of the model. Standardised residuals are the residuals divided by an estimate of their standard deviation. They are used because it is difficult to estimate a

universal cut-off point for what constitutes a large residual that stands out with unstandardized residuals (Field, 2009). They seem to be relatively normally distributed across the regression line and there seem to be no residual with an absolute value of 3 or more. Thus, the analysis seems to meet the assumption of normality.

Q-Q Plot

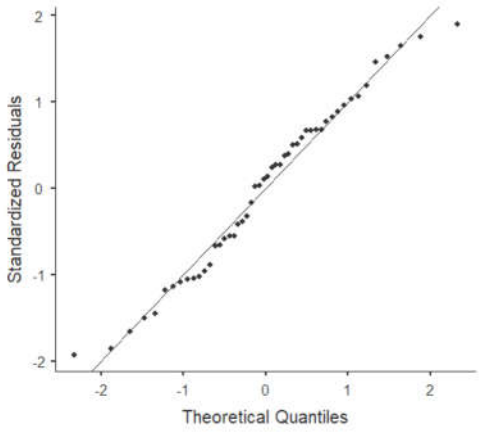


Figure 4.6 Standardised residuals' plot

Table 4.6 shows the results of the hierarchical regression analysis. Analysis shows that R^2 is 0.512. Squared multiple correlation (R^2) is the proportion of variation in the outcome variable that is predicted from the best linear combination of the predictor variables (Tabachnick & Fidell, 2013). In other words, R^2 is the amount of variance in the outcome variable that can be accounted for by the model (Field, 2009), so in this case 51.2% of the variance in reading comprehension in English can be explained by the predictor variables of the model. This means that the remaining 48.8% is due to other factors. ΔR^2, which is also called R^2 change, in Table 4.7 shows the contribution of the second block of predictor variables to explaining the variance in the outcome variable (Field, 2009). After the addition of the second block of predictors the value of ΔR^2=0.148. Thus, 14.8% of the variance in reading comprehension in English is explained by vocabulary knowledge in English.

Model Fit Measures

				Overall Model Test			
Model	**R**	**R^2**	**Adjusted R^2**	**F**	**df1**	**df2**	**p**
1	0.603	0.364	0.350	27.4	1	48	< .001
2	0.715	0.512	0.480	16.1	3	46	< .001

Table 4.6 Model fit measures

Model Comparisons

Comparison							
Model		**Model**	**ΔR^2**	**F**	**df1**	**df2**	**p**
1	-	2	0.148	6.98	2	46	0.002

Table 4.7 Model comparisons

Table 4.8 shows the model's coefficients when all the predictor variables are included. Reading comprehension in L1 ($p<.001$) and vocabulary size ($p=0.014$) have a statistically significant p value. The second vocabulary variable, LexTale, has larger p-value ($p=0.469$). However, this does not mean that this variable does not contribute to prediction of reading comprehension in English. It is possible for a predictor variable to seem unimportant when it is actually highly correlated with the outcome variable. This may happen if the area of that correlation is whittled away by other predictor variables. In this case the unique contribution of the predictor variable is small despite the correlation with the outcome variable (Tabachnick & Fidell, 2013). As sheen in Table 4.4, Spearman's ρ for LexTale is 0.371. This predictor variable has a relatively strong correlation with the outcome variable which shows

that it is collectively important for the model but it does not appear to be important in its unique contribution in the regression model with p=0.469.

Model Coefficients – Reading comprehension in L2

Predictor	Estimate	SE	t	p
Intercept	-0.108	0.148	-0.730	0.469
Reading comprehension in L1	0.766	0.177	4.317	< .001
Vocabulary size	0.003	0.001	2.559	0.014
LexTale	0.001	0.002	0.729	0.469

Table 4.8 Model coefficients

5. Discussion

As mentioned in Chapter 3, validity and reliability are the most important prerequisites of sufficient quality in scientific research. Whether conclusions can be drawn from a research, depends on whether the research is valid and reliable. Whether these requirements have been met in this study will be discussed in this chapter. But in the first part of the chapter, the results of the study will be discussed in light of the theoretical background presented in Chapter 2.

5.1 Study's results

The theoretical background and previous research have shown that both L2 vocabulary knowledge and L1 reading comprehension are important predictors of L2 reading comprehension and that there is cross-linguistic transfer between L1 and L2. There is not yet an agreement though, on whether this transfer occurs from the beginning or there has to be a certain level of L2 proficiency acquired before the transfer occurs. In this section the results of the study will be discussed in light of the earlier presented theory. First the relationship between L2 reading comprehension with L2 vocabulary knowledge and L1 reading comprehension will be discussed. Then, to what extend the results of the hierarchical regression analysis support the existing theory about cross-linguistic transfer will be discussed.

The relationship between L2 reading comprehension, L2 vocabulary knowledge and L1 reading comprehension

The results of this study have shown that Norwegians have a not as high level of proficiency in English, as it was supported by Brevik et al. (2016). L2 reading comprehension test had a mean score of 0.66, vocabulary size test 51.8 and the lexical decision test 75.6. This result means that they do have a good level of English proficiency, though not very high. However, this is in line with the Lexical Quality Hypothesis (Perfetti & Hart, 2002) which claims that reading skill is supported by the reader's high quality of a word representation which means its orthography, phonology and semantics. So, the good performance in the L2 reading comprehension test is supported by the good performance in the two L2 vocabulary tests.

The correlation of L2 reading comprehension and L2 vocabulary knowledge was high. The correlation between L2 reading comprehension and the vocabulary size test was ρ=0.51 and between L2 reading comprehension and LexTale ρ=0.37, which is a strong enough correlation because in educational research it cannot be expected to have higher correlations than between 0.20 and 0.40 between variables (Gall et al, 2007). This finding is in line with the findings of many studies that support that L2 vocabulary plays an important role in L2 reading comprehension (Raudszus et al., 2019; Yamashita & Shiotsu, 2017; O'Connor et al., 2019; Schaars et al., 2019; Verhoeven, Voeten & Vermeer, 2019), especially among adults (Jeon & Yamashita, 2014; Raudszus et al., 2018). L1 reading comprehension correlates strongly with L2 reading comprehension (ρ=0.56) as well. This is in line with other research findings about their relationship that support that the two variables correlate strongly (Brevik et al., 2016).

L2 vocabulary knowledge accounted for 14.8% of the L2 reading comprehension which also is in line with the Reading Systems Framework (Perfetti & Stafura, 2014) which supports that lexical knowledge is a prime predictor of reading comprehension because of its critical role in word-to-text integration process.

Results about cross-linguistic transfer

According to the Linguistic Interdependence Hypothesis (Cummins, 1979), proficiency acquired in L1 can transfer or facilitate learning in L2. L1 and L2 are interdependent and depend on common underlying proficiency, so the hypothesis is that skilled readers in L1 will also read well in L2. This study's findings support that hypothesis. Performance in both L1 and L2 reading comprehension tests was very good (L1 reading comprehension test had a mean score of 0.72 and L2 reading comprehension test 0.66) and L1 reading comprehension accounted for 36.4% of the variance of L2 reading comprehension. Thus, the findings suggest that the participants were skilled in both L1 and L2 reading and there was transfer of skills from L1 to L2.

However, the findings seem to also support the Linguistic Threshold Hypothesis (Cummins, 1979) which states that L1 skills transfer to L2 only when a certain level of L2 language proficiency has been acquired. In this study L2 vocabulary knowledge accounted for 14.8% of L2 reading comprehension. Thus, there was transfer of skills between L1 and L2 reading comprehension while L2 vocabulary knowledge was high (mean score of 51.8 and 75.6) and

played a significant role in the prediction of L2 reading comprehension. This is not enough evidence though, to conclude that the transfer occurs due to the high level of L2 proficiency. It just supports the hypothesis that when there is such a level of L2 proficiency, L1 skills are transferred to L2. Thus, the study's results do not conclude on whether transfer between the L1 and L2 reading comprehension occurs anyway or whether a good level of L2 language proficiency should first be acquired.

Furthermore, these findings seem to support Alderson's (1984 in Yamashita & Shiotsu, 2017) and Jeon's and Yamashita's (2014) answer in the question of whether L2 reading is a reading problem or a language problem. The question is about whether L2 reading difficulties are better explained by limited L2 knowledge or by weak L1 reading ability. So, whether L2 knowledge or L1 reading ability is more important in L2 reading development. They suggest that it seems to be a matter of language, although it is still significantly a reading matter. In this study L2 reading comprehension was explained by both L1 reading comprehension and knowledge of L2 vocabulary which together accounted for 51.2% of the total variance. Thus, it seems to be both a language and a reading matter.

Finally, Bernhardt's (2011) compensatory theory states that if the reader has a weakness in one area, it will be compensated for by knowledge from another area because of the transfer between L1 and L2. According to her model, L2 language knowledge accounts for up to 30% of L2 reading comprehension and L1 literacy for up to 20%. In this study L2 vocabulary knowledge accounts for 14.8% of L2 reading comprehension. This is quite in line with Bernhardt's model because vocabulary knowledge is just a part of L2 language knowledge and the remaining part could possibly account for the rest 15.2% of the proposed 30%. L1 literacy, according to her, accounts for up to 20% of L2 reading comprehension, but in this study, L1 reading comprehension, which is included in L1 literacy, accounts for 36.4%. This may be due to the close distance of the two languages. As mentioned earlier, Norwegian and English are both alphabetic languages of the same origin, they are Germanic languages and share a lot of cognates.

5.2 Validity and reliability

Construct validity

Construct validity refers to whether the measures that were used in the study really measure the constructs they claim to measure (Gall et al., 2007). In this study construct validity refers to whether the different tests that were used measure the concepts of L1 reading comprehension, L2 vocabulary knowledge and L2 reading comprehension as they were defined in the theoretical part. A potential threat of construct validity comes from the fact that the data that were used were part of a very large pool of data from a research about eye-movements during reading in L1 and L2. The data collected for Norwegian as the L1 included measures related to different aspects of language knowledge and reading ability in both L1 and L2. Thus, there is a limitation in how thorough all these various variables were examined, because of the limited time available for each. Specifically, only one measure of L1 reading comprehension and one of L2 reading comprehension were used. For L2 vocabulary knowledge two measures were used.

A strength of this study might be the type of measures used. As mentioned earlier, it is very difficult to measure reading comprehension because it depends on many characteristics, such as the type of text, the type of questions and the reader's prior knowledge. Despite the fact that all measures are tapping the same underlying construct, different measures assess different types of information about reading comprehension and have different demands on language knowledge (Cain, 2010). In this study yes/no sentence judgements and multiple-choice response format were used. Both formats probe the reader's memory and have relatively low processing demands, since they do not need complex verbal response. These formats are the most frequently used when it comes to reading comprehension. For vocabulary knowledge a receptive vocabulary test and a lexical decision test, which also depends on receptive vocabulary, were used. Reading comprehension is a receptive skill, so receptive vocabulary tests might be more suitable for this study, since it has been found to be critical for adequate reading comprehension and scores from a receptive test will correlate more strongly with scores from reading comprehension tests (Jeon & Yamashita, 2014; Verhoeven et al., 2019).

Another potential threat of construct validity concerns cross-cultural validity. The measure of L1 reading comprehension was developed by the master students working with the project, since the texts and the questions were either translated or adapted from the original English

ones. No pilot test for this measure was done, due to the limited time available by the study program. However, face validity of the test was good, since it was approved by several Norwegian native speakers, who did not participate in the study, as a good measure for reading comprehension of these texts.

Another threat to construct validity is the so-called "task-impurity" problem. This means that tests do not only measure the variables they were supposed to measure, but other variables as well, since all tasks require multiple skills to be solved (Miyake et al., 2000). This problem is present when it comes to measuring reading comprehension and vocabulary knowledge. Measuring reading comprehension alone is not possible since reading comprehension depends on other skills such as decoding, vocabulary, grammar, syntax, inference making and monitoring strategies. Vocabulary knowledge also depends both in the knowledge of the meaning of single words as well as utterances, idioms and expressions and how individual words relate and associate with each other. The two measures of L2 vocabulary knowledge in this study used only individual words to measure it, so there was no need for knowledge of idioms, expressions, utterances and relations between words.

Statistical conclusion validity

Statistical conclusion validity is about whether valid conclusions can be drawn by the results of statistical analysis about the relationship between variables. This implies that the relationship between the variables is statistically significant and strong enough to be of theoretical significance (Lund, 2002). In this study statistically significant correlations where found between all variables. If a relationship is found between variables, the possibility to commit a Type I error should be considered. Type I error is concluding that there is relationship between the variables, when actually there is not. A significance level of 0.05 is usually used (Field, 2009). A significance level of 0.05 means that if the collection of data is replicated 100 times, in 5 cases it can be expected to be found a significant relationship between the variables, when actually there is not (Field, 2009). In this study some correlations were significant at the 0.05 level, while others at the 0.01 level. Thus, it may be concluded with reasonable certainty that relationships between variables are significant and that there is a relatively small likelihood of having committed a Type I error. A Type II error is concluding that there is not relationship between variables, even though there is (Field,

2009). In this study the relationship between all variables was found to be of statistical significance, so a Type II error would not be possible to occur.

In case that a statistically significant relationship is found between variables, the strength of the relationship should also be assessed (Lund, 2002). Within every research field, the strength a relationship should have to be considered of theoretical significance varies. Educational research involves research of complex phenomena and as a result it cannot be expected to have higher correlations than between 0.20 and 0.40 between variables (Gall et al, 2007). In this study the correlations were of different sizes. L2 reading comprehension correlated higher than 0.40 with L1 reading comprehension (ρ=0.56) and with the vocabulary size measure (ρ=0,51) and had a strong enough correlation with the lexical decision measure (ρ=0.37). Correlations between the independent variables were also strong. L1 reading comprehension correlated with the vocabulary measure with ρ=0.37 and with the lexical decision measure also with ρ=0.37. There is little doubt this high correlations have theoretical significance.

Internal validity

Internal validity is about the extent to which safe conclusions about the relationships between the variables can be drawn. There is no definitive evidence of causality in a non-experimental study. The statistical relationship will always be compatible to several possible causal relationships (Kleven, 2002). This study is a prediction study which investigates whether the scores on one variable can be predicted from the scores of another variable. This can be done by looking at the correlation between the independent variables and the dependent variable and the size of common variance. The goal would preferably be to investigate whether L1 reading comprehension and L2 vocabulary knowledge influence L2 reading comprehension. However, this implies causation. Since this study is non-experimental, there cannot be any experimental control and because of that non-experimental studies have lower internal validity than experimental ones (Kleven, 2002).

Thus, it is not possible to draw valid causal conclusions because there are several possible underlying causal relationships. It is not possible to conclude with certainty that there are not any other factors that may explain the relationship. This is called the “third-variable problem” and means that another variable could be responsible for the relationship (Kleven, 2002). If a Norwegian speaking adult has good L2 reading comprehension skill, could be due to other

factors as well. This is especially true for research of complex phenomena, such as educational research, because a researcher can never be completely sure on including all factors that can possibly be influencing (Gall et al., 2007). Furthermore, it is not possible to draw certain conclusions about the direction of the relationship (Kleven, 2002). Does L1 reading comprehension affect L2 reading comprehension? Does L2 reading comprehension affect L1 reading comprehension? Is there a mutual influence? It is not possible to conclude on any of these assumptions.

External validity

Good external validity is achieved when the relationship between variables can be generalised to relevant persons, situations and times (Lund, 2002). In order to generalise, the study's sample should meet several requirements. It is crucial that the sample is representative of the population. The population of this study is people that have Norwegian as L1 and English as L2 with various socioeconomic statuses and levels of education. This includes the majority of Norwegians that went to a Norwegian school, since Norwegians are taught English as a compulsory subject from the first until the eleventh year of the school and the level of English proficiency among Norwegians has long been high (Brevik et al., 2016). The sample of the study was adult persons under the age of 30 that had Norwegian as their mother tongue and were proficient in English. The sample had different socioeconomic status, different number of years of education and different fields of education and work and did not have any diagnosed learning, audio or vision difficulties.

A randomised sampling would be the most appropriate for representativeness. Randomised sampling is when all individual in the targeted population have equal and independent chance of being selected as the study's sample (Gall et al., 2007). This type of sampling however, is not easy to achieve because it requires a lot of resources. So, for practical reasons, sampling was made informally among the acquaintances of the master students that are working with the project. Thus, the sample is probably not representative of the whole Norwegian population but hopefully of a sufficiently wide range of it which is the population of young adult Norwegians that live, study or work in big cities.

The size of the sample is also linked to external validity and the generalisability of the study's results. The bigger the size of the sample, the more stable the estimates become and the more the uncertainty about the variables decreases (Field, 2009). In a correlational research, such

as this one, a minimum of 30 participants is desirable (Gall et al., 2007). The sample of the present study consists of 50 participants, which is a small sample.

Another concern about the generalisability of the results is not about the sample of people participating in the study, rather about the sample of texts used to measure reading comprehension in both L1 and L2. The question is whether the results of the study can be generalised in reading comprehension of other types of texts as well. The texts used were short texts, varying from 6 to approximately 16 lines each. The complexity of the texts varied from elementary level to college level texts. The entries were taken from Wikipedia for L1 and from a larger reading comprehension test for L2. There were various topics about culture, technology, health and nature some of which might be quite familiar to the participants while others might be totally unknown. Since the texts that people come across in their life vary in length, complexity, source, familiarity and topic, the texts used for reading comprehension in this study appear to be similar to a range of brief texts of low to medium complexity.

Reliability

Reliability is the degree to which an assessment tool produces stable and consistent results. In quantitative research reliability refers to the degree to which research is accurate and credible (Cozby, 2015). Measurement error can be a potential threat of the study's reliability. There many types of errors that may occur during measurement. For example, the tests could be administered and scored in different ways by the different administrators or there could be random errors while administering and scoring them. Also, participants could not be motivated enough to take the test or they might have a bad day, not feeling well or not having slept enough and being tired. Furthermore, the room might not be suitable, the equipment might have some problem, there might be interruptions or noise or the environment might not be friendly.

Several measures were taken to limit the possibility of measurement errors. All three administrators had a very thorough training about the whole process of the testing and each test by more than one person and clear instructions about administrating and scoring the tests were given to them. All had a relevant background from psychology, education and special needs education. All of the tests were performed on the computer, which simplified administration and scoring and limited scoring errors. Participants were encouraged to make themselves comfortable and take as many breaks as they needed, at least one. There were

cases that some participants were encouraged to stop the process because they were not feeling well. The room was quiet, with all the needed equipment and there were no interruptions. The administrators tried to make the environment friendly by chatting with the participants before the testing and during the breaks.

Reliability analysis showed that not all variables had high reliability. As mentioned in Chaper 3, 0.80 and 0.70 have been proposed as the limit values to provide reliable results (Gall et al., 2007). The tests concerning L2 vocabulary knowledge had a high reliability. The vocabulary size test had a reliability of α=0.98 and the lexical decision test a split-half reliability of 0.81. However, the tests concerning reading comprehension had a relatively moderate reliability. L2 reading comprehension test had a reliability of α=0.68 and L1 reading comprehension test α=0.62. According to Kline (1999), a reliability below 0.70 might be expected in tests if the phenomenon being measured is complex. As mentioned earlier, measuring reading comprehension is very difficult because it is a complex skill that includes a lot of other skills. Furthermore, the test of L1 reading comprehension was developed by the students working with the project by translating or adapting the original English one, and thus, it is not unexpected that its reliability would be a little low.

Summary reflections

Hierarchical regression analysis has shown that L1 reading comprehension was the best predictor of L2 reading comprehension, accounting for 36.4% of the variance. The two variables correlated strongly as it is supported by literature (Brevik et al., 2016). L2 vocabulary knowledge was as well an important predictor of L2 reading comprehension explaining 14.8% of the variance, supporting the existing literature (Raudszus et al., 2019; Yamashita & Shiotsu, 2017; O'Connor et al., 2019; Schaars et al., 2019; Verhoeven, Voeten & Vermeer, 2019). Norwegians were skilled readers in both L1 and L2 and had a good level of English proficiency. Cross-linguistic transfer occurred between L1 and L2 as it is shown by the variance that is accounted for by L1 reading comprehension, supporting the Linguistic Interdependence Hypothesis (Cummins, 1979). However, the contribution of L2 vocabulary knowledge was significant supporting the Linguistic Threshold Hypothesis (Cummins, 1979) and the Compensatory Theory (Bernhardt, 2011). No certain conclusion could be drawn about when the transfer occurs.

6. Conclusion

The purpose of this study has been to illustrate the relationship between L2 reading comprehension, L2 vocabulary knowledge and L1 reading comprehension. The study is based on the following question: *To what extent does L2 (English) vocabulary account for L2 reading comprehension, after accounting for L1 (Norwegian) reading comprehension, in adult native Norwegian speakers?* The goal was to gain insight into the relationship between the knowledge of English vocabulary and reading comprehension in English as a second language as well as whether there was cross-linguistic transfer between reading comprehension in Norwegian as the first language and English as the second.

The mean scores of vocabulary and reading comprehension tests in L2 show that in this study the level of English knowledge for Norwegians was good but lower than expected from other studies' results. Nonetheless, findings support the Lexical Quality Hypothesis since good knowledge of vocabulary supports good reading comprehension. Correlation analysis shows that there was a high correlation between L2 vocabulary and L2 reading comprehension supporting the findings of prior studies that L2 vocabulary plays a significant role in L2 reading comprehension. The findings of the hierarchical regression analysis about the percentage of L2 reading comprehension variance explained by L2 vocabulary knowledge support the Reading Systems Framework which as well suggests that lexical knowledge is a prime predictor of reading comprehension.

This study's findings also support transfer between L1 and L2. L1 reading comprehension correlated strongly with L2 reading comprehension and a large percentage of L2 reading comprehension variance was explained by L1 reading comprehension. The results support both the Linguistic Interdependence Hypothesis and the Linguistic Threshold Hypothesis, since there was cross-linguistic transfer between L1 and L2 and there was a good level of L2 knowledge. However, there is not enough evidence to indicate whether transfer occurs after the acquisition of a certain level of L2, or it occurs from the moment a person begins to learn the second language. In addition, the study's results support that L2 reading comprehension is a matter of both language and reading, since L2 vocabulary knowledge and L1 reading comprehension explained together more than the half of the variance of L2 reading comprehension. The findings were also in line with the claims of the Compensatory Theory about the contribution of L2 language knowledge and L1 literacy in L2 reading, though in

this study L1 might have played a more significant role in L2 than suggested by the model, because of the proximity of the two languages in focus.

Although cross-linguistic transfer has been a topic of interest for many studies, a lot of research is still required for this phenomenon to be fully understood. There should be more research about the transfer between L1 and L2 using different languages with different levels of similarity between them and different orthographies. Also, there should be research with other languages than English as L2, which is the dominant L2 in most existing research. More research should as well be conducted for determining when the transfer occurs, rather than whether transfer occurs. Further research is also required to uncover other variables that contribute to the remaining half of L2 reading comprehension's variance. Regarding Norwegian speakers, more research is required about the contribution of Norwegian as L1 and other languages apart from English as L2. Finally, the level of transfer between L1 and L2 in children that are not yet skilled readers and have not yet acquired a good L2 knowledge should be researched.

References

Bernhardt, E. B. (2011). *Understanding Advanced Second – Language Reading*. New York: Routledge.

Bernhardt, E.B. & Kamil, M.L. (1995). Interpreting Relations between L1 and L2 Reading: Consolidating the Linguistic Threshold and the Linguistic Interdependence Hypotheses. *Applied Linguistics*. 16(1), 15-34.

Brevik, L.M., Olsen, R.V. and Hellekjær, G.O. (2016). The complexity of second language reading: Investigating the L1-L2 relationship. *Reading in a Foreign Language*. 28(2), 161-182.

Cain, K. (2010). *Reading Development and Difficulties*. UK: BPS Blackwell.

Castles, A., Rastle, K., Nation, K., (2018). Ending the Reading Wars: Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest.* 19 (1), 5-51.

Chung, S.C., Chen, X. & Geva, E. (2019). Deconstructing and reconstructing cross-language transfer in bilingual reading development: An interactive framework. *Journal of Neurolinguistics*. 50, 149-161.

Clarke, M.A. (1978). Reading in Spanish and English: Evidence from adult ESL students. *Language Learning*. 29(1), 121-150.

Consensus-based Standards for the selection of health Measurement Instruments (COSMIN). (2018). Retrieved from: https://www.cosmin.nl/tools/cosmin-taxonomy-measurement-properties/

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experementation. Design & Analysis Issues for Field Settings*. Boston: Houghton Mifflin Company.

Cop, U., Dirix, N., Drieghe, D., Duyck, W. (2017). Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior Research Methods*. 49, 602-615. Cozby, P. C. (2015). *Methods in Behavioral Research*. (12th ed.). Ohio: McGraw-Hill.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16 (3), 297-334.

Cumming, G., Calin-Jageman, R. (2017). *Introduction to the new statistics*. New York: Routledge.

Cummins, J. (1979). Linguistic Interdependence and the Educational Development of Bilingual Children. *Review of Educational Research*. 49(2), 222-251.

de Groot, A.M.B. (2006). Effects of Stimulus Characteristics and Background Music on Foreign Language Vocabulary Learning and Forgetting. *Language Learning*. 56(3), 463-506.

Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders.* Baltimore: Paul H. Brookes.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *The British Picture Vocabulary Scale* (2nd ed.). London: nferNelson.

Field, A. (2009). *Discovering statistics using SPSS: (and sex and drugs and rock 'n' roll)* (4. utg.). Los Angeles: SAGE.

Field, A. (2016). *An adventure in statistics: The reality enigma*. London: SAGE.

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: an introduction*. Boston: Allyn and Bacon.

Gough, P.B. & Tunmer, W.E. (1986). Decoding, Reading and Reading Disability. *Remedian and Special Education*. 7, 6-10.

Hjetland, H.N., Brinchmann, E.I., Scherer, R. & Melby-Lervåg, M. (2017). Preschool predictors of later reading comprehension ability: a systematic review. *Campbell Systematic Reviews*. 14.

Holmqvist, K., Nystrom, M, Anderson, R., Dewhurst, R., Jarodzka, H., van de Weijer, J. (2011). *Eye Tracking A Comprehensive Guide to Methods and Measures.* Oxford: University Press.

Jeon, E.H. & Yamashita, J. (2014). L2 Reading Comprehension and Its Correlates: A Meta-Analysis. *Language Learning*. 64(1), 160-212.

Jiang, X. (2011). The Role of First Language Literacy and Second Language Proficiency in Second Language Reading Comprehension. *The Reading Matrix*. 11(2), 177-190.

Kleven, T. A. (2002). Ikke-eksperimentelle design. In Lund (Ed.), *Innføring I forskningsmetodologi* (pp. 265-286). Oslo: Unipub.

Lee., J.W. & Schallert, D.L. (1997). The Relative Contribution of Language Proficiency and L1 Reading Ability to L2 Reading Perfomance: A Test of the Threshold Hypothesis in an EFL Context. *Tesol Quarterly*. 31(4), 713-739.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior research methods*, *44*(2), 325-343.

Lervåg, A. & Aukrust, V.G. (2010). Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *Journal of Child Psychology and Psychiatry*. 51(5), 612-620.

Li, P., Clariana, R. B. (2019). Reading comprehension in L1 and L2: An integrative approach. *Journal of Neurolinguistics*. 50, 94-105.

Lund, T. (2002). Metodologiske prinsipper og referanserammer. In Lund (Ed.), *Innføring i forskningsmetodologi* (s. 79-123). Oslo: Unipub.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*.

Melby-Lervåg, M. & Lervåg, A. 2011. Cross-linguistic transfer of oral language, decoding, phonological awareness, and reading comprehension: A meta-analysis of the correlational evidence. *Journal of Research in Reading*. 34(1), 114-135.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A. & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cognitive Psychology, 41*(1), 49-100.

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

O'Connor, M., Geva, E. & Koh, P.W. (2019). Examining Reading Comprehension Profiles of Grade 5 Monolinguals and English Language Learners Through the Lexical Quality Hypothesis Lens. *Journal of Learning Disabilities*. 52(3), 232-246.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of Functional Literacy*, 11, 67–86.

Perfetti, C. & Stafura, J. (2014). Word Knowledge in a Theory of Reading Comprehension. *Scientific Studies of Reading*. 18(1), 22-37.

Raudszus, H., Segers, E., Verhoeven, L. (2018). Lexical quality and executive control predict children's first and second language reading comprehension. *Reading and Writing: An Interdisciplinary Journal*. 31(2), 405-424.

Raudszus, H., Segers, E., Verhoeven, L. (2019). Situation model building ability uniquely predicts first and second language reading comprehension. *Journal of Neurolinguistics*. 50, 106-119.

Rayner, K., Schotter, E., Masson, M., Potter, M., Treiman, R., (2016). So Much to Read, So Little Time: How Do We Read, and Can Speed Reading Help? *Psychological Science in the Public Interest.* 17 (1), 4-34.

Saiegh-Haddad, E. & Geva, E. (2010). Acquiring reading in two languages: an introduction to the special issue. *Reading and Writing: An Interdisciplinary Journal*. 23, 263-267.

Schaars, M.M.H, Segers, E. & Verhoeven, L. (2019). Cognitive and linguistic precursors of early first and second language reading development. *Learning and Individual Differences*. 72, 1-14.

Seymour, P.H.K., Aro, M. & Erskine, M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*. 94, 143-174.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson.

The Norwegian National Research Ethics Committees (NESH). (2016). Retrieved from: https://www.etikkom.no/en/ethical-guidelines-for-research/guidelines-for-research-ethics-in-the-social-sciences--humanities-law-and-theology/

van den Bosch, L.J., Segers, E. & Verhoeven, L. (2019). The Role of Linguistic Diversity in the Prediction of Early Reading Comprehension: A Quantile Regression Approach. *Scientific Studies of Reading*. 23(3), 203-219.

Vehoeven, L., Perfetti, C. & Pugh, K. (2019). Cross-linguistic perspectives on second language reading. *Journal of Neurolinguistics*. 50, 1-6.

Verhoeven, L., Voeten, M. & Vermeer, A. (2019). Beyond the simple view of early first and second language reading: The impact of lexical quality. *Journal of Neurolinguistics*. 50, 28-36.

Yaghoub Zadeh, Z., Farnia, F. & Geva, E. (2012). Toward modelling reading comprehension and reading fluency in English language learners. *Reading and Writing: An Interdisciplinary Journal.* 25, 163-187.

Yamashita, J. (2002). Mutual compensation between L1 reading ability and L2 language proficiency in L2 reading comprehension. *Journal of Research in Reading*. 25(1), 81-95.

Yamashita, J., & Shiotsu, T. (2017). Comprehension and Knowledge Components That Predict L2 Reading: A Latent- Trait Approach. *Applied Linguistics*. 38(1), 43-67.

Appendix 1: Norwegian reading comprehension test

Prototype English version

1) [TRANSLATE] In ancient Roman religion and myth, Janus is the god of beginnings and gates. He has a double nature and is usually depicted as having two faces, since he looks to the future and to the past. Janus presided over the beginning and ending of conflict, and hence war and peace. The doors of his temple were open in times of war and closed during peacetime. As the god of gates, he was also associated with entering and exiting the doors of homes. Janus frequently symbolized change and transitions, such as the progress from one condition to another, from one vision to another, and the growth of young people into adulthood. Hence, Janus was worshipped at the beginning of the harvest and planting times, as well as at marriages, deaths, and other beginnings. Janus had no specialized priest assigned to him, but the high priest himself carried out his ceremonies. Janus represented the middle ground between barbarism and civilization, rural and urban space, youth and adulthood. The ancient Greeks had no equivalent to Janus, whom the Romans claimed as distinctively their own.

1. Was Janus a god of doors and passages?
2. Does Janus symbolize the night time?
3. Were the doors of Janus' temple open in time of peace?
4. Is Janus a Greek god?

2) The shaka sign, sometimes known as "hang loose", is a gesture of friendly intent often associated with Hawaii and surf culture. The gesture is made by extending the thumb and smallest finger while holding the three middle fingers curled and gesturing in salutation while presenting the front or back of the hand; the hand may be rotated back and forth for emphasis. The shaka sign was adopted from local Hawaiian culture and customs by visiting surfers in the 1960s, and its use has spread around the world. It is primarily used as a greeting gesture or to express thanks, acknowledgement, or even praise from one individual to another. Residents of Hawaii use the shaka to convey the "Aloha Spirit": a concept of friendship, understanding, compassion, and solidarity among the various ethnic cultures that reside in Hawaii. While the exact origin of the sign is unclear, some have suggested its origin was derived from Spanish immigrants, who folded their middle fingers and took their thumbs to their lips as a friendly gesture to represent sharing a drink with the natives they met in Hawaii.

1. Did Spanish immigrants adopt the shaka sign from Hawaiian residents?
2. Does the "Aloha Spirit" convey positive interpersonal relations?
3. Is the shaka sign used for teasing?
4. Has the shaka sign spread around the world in the 20th century?

3) [TRANSLATE] In competitive sports, doping is the use of banned performance-enhancing drugs by athletic competitors. The term doping is widely used by organizations which regulate sporting competitions. The use of drugs to enhance performance is largely considered unethical, and is therefore prohibited by most international sports organizations, including the International Olympic Committee. Furthermore, athletes who take explicit measures to evade detection exacerbate the ethical violation with overt deception and cheating. Despite its prevalence in the headlines recently, doping is not a new phenomenon—in fact, it is as old as sport itself. From the use of substances in ancient chariot races to more recent controversies in baseball and cycling, popular views among athletes have varied widely over the years. In recent decades, authorities and sporting organizations have tried to strictly regulate the use of drugs in sport. The primary reasons for this ban are the health risks of performance-enhancing drugs, the equality of opportunity for athletes, and the positive example to the public set by drug-free sport. Anti-doping authorities have repeatedly emphasized that using performance-enhancing drugs goes against the "spirit of sport".

1. Can doping have adverse health effects?
2. Do athletes use doping to calm down before their competition?
3. Is doping an issue only in specific disciplines?
4. Was the use of doping first observed in ancient wrestling competitions?

4) The thylacine was the largest known carnivorous marsupial of modern times. It is commonly known as the Tasmanian tiger (because of its striped lower back) or the Tasmanian wolf (because of its canine-like appearance, traits and attributes). Native to continental Australia, Tasmania, and New Guinea, it is believed to have become extinct in the twentieth century. The thylacine was one of only two marsupials to have a pouch in both sexes. The male had a pouch that protected the external reproductive organs while running through thick brush. The thylacine has been described as a formidable predator because of its ability to survive and hunt prey in extremely sparsely populated areas. The thylacine had become extremely rare or extinct on the Australian mainland before British settlement of the continent, but it survived on the island of Tasmania. Intensive hunting encouraged by bounties is generally blamed for its extinction, but other contributing factors may have included disease, the introduction of dogs, and human encroachment into its habitat. Despite its official classification as extinct, sightings are still reported, though none have been conclusively proven.

1. Do male thylacines have a pouch?
2. Did introduction of sheep contribute to extinction of thylacine?
3. Is the thylacine compared to a tiger because of its tail?
4. Did the thylacine live in Asia?

5) World Environment Day (WED) is celebrated on the fifth of June every year and is the United Nation's principal vehicle for encouraging awareness and action for the protection of the environment. First held in 1974, it has been a flagship campaign for raising awareness on emerging environmental issues from marine pollution, human overpopulation, and global warming. WED has grown to become a global platform for public outreach, with participation from over one hundred countries annually. Each year, WED chooses a new theme that major corporations, non-governmental organizations, communities, governments and celebrities worldwide adopt to advocate environmental causes. For example, the theme for 2018 is "Beat Plastic Pollution". The goal is to encourage people to change their everyday habits in ways which can reduce plastic pollution. Specifically, the campaign hopes to reduce the prevalence of single-use or disposable plastic items, as they have severe environmental consequences. In response to this campaign, the Indian government has pledged to eliminate all single-use plastic in India by 2022.

1. Is the WED of 2018 focused on India eliminating plastic use?
2. Do governments get involved in the advocating of WED?
3. Is WED celebrated on a different date every year?
4. Is overpopulation of Earth one of common WED themes?

6) A monocle is a type of corrective lens used to correct or enhance the vision in only one eye. It consists of a circular lens, generally with a wire ring around the circumference that can be attached to a string or wire. The other end of the string is then connected to the wearer's clothing to avoid losing the monocle. During the late nineteenth century, the monocle was generally associated with wealthy, upper-class men. Combined with a long coat and a top hat, the monocle completed the costume of the stereotypical late nineteenth century capitalist. Monocles were also accessories of military officers from this period. Despite their prevalence in the late nineteenth century, monocles are rarely worn today. This is due in large part to advances in optometry, which allow for better measurement of refractive error, so that glasses and contact lenses can be prescribed with different strengths in each eye.

1. Were monocles often worn on a string?
2. Was the monocle associated with a stereotypical look?
3. Were monocles deemed obsolete because they went out of style?
4. Were monocles worn by the members of social and military elite?

7) [TRANSLATE] Wine tasting is the sensory examination and evaluation of wine. While the practice of wine tasting is as ancient as its production, a more formalized methodology has

slowly become established from the late middle ages onwards. Modern, professional wine tasters use a constantly evolving specialized terminology which is used to describe the range of perceived flavors, aromas and general characteristics of a wine. In recent years, results challenging the reliability of wine tasting in both experts and consumers have surfaced. For example, studies showed that people expect more expensive wine to have more desirable characteristics than less expensive wine: When tasters are given wine that they are falsely told is expensive they virtually always report it as tasting better than the very same wine when they are told that it is inexpensive. Other studies show that tasters' judgment can be prejudiced by knowing details of a wine, such as geographic origin, reputation, or other considerations. Objective wine tasting therefore requires the wine to be served blind – that is, without the taster having seen the label or bottle shape. Blind tasting may also involve serving the wine from a black wine glass to mask the color of the wine.

1. Are wine tasters reliable in their judgments of wine quality?
2. Are experienced wine tasters less prejudiced towards wine than a lay person?
3. Does blind-tasting ensure wine tasting to be less subjective?
4. Is wine served in a black glass to mask its color?

8) Orange juice is a liquid extract of the orange tree fruit, produced by squeezing oranges. Commercial orange juice with a long shelf life is made by pasteurizing the juice and removing the oxygen from it. This removes much of the taste, necessitating the later addition of a flavor pack, generally made from orange products. Additionally, some juice is further processed by drying and later rehydrating the juice, or by concentrating the juice and later adding water to the concentrate. The health value of orange juice is debatable: it has a high concentration of vitamin C, but also a very high concentration of simple sugars, comparable to soft drinks. As a result, some government nutritional advice has been adjusted to encourage substitution of orange juice with raw fruit, which is digested more slowly, and limit daily consumption.

1. Are there crucial health benefits to drinking orange juice?
2. Is commercial orange juice pasteurized in order to taste better?
3. Do companies remove oxygen from orange juice?
4. Is orange juice similar to soda in its sugar content?

9) Beekeeping is the maintenance of bee colonies by humans, typically with the use of man-made hives. The species most commonly used in bee colonies are honey bees. A beekeeper maintains these colonies in order to collect the honey and other products that the hive produces (such as beeswax), but also to pollinate crops, or to produce bees for sale to other beekeepers. The domestication of bees is depicted in Egyptian art from over four thousand years ago. Back then, simple hives and smoke were used for beekeeping, and honey was stored in jars – some

of which were found in the tombs of pharaohs. This process was largely unchanged until the eighteenth century, when an improved understanding of the biology of bees led to the invention of the moveable comb hive in Europe, allowing honey to be harvested without destroying the entire colony. Today, some beekeepers believe that the more stings a beekeeper receives, the less irritation each one causes, and it is considered important for the safety of the beekeeper to be stung a few times a season. Indeed, recent studies found that beekeepers have high levels of antibodies that react to the major antigen of bee venom.

1. Was beekeeping practiced in ancient Egypt?
2. Is beekeeping widely obsolete today?
3. Can beekeepers build up immunity to bee venom?
4. Is beekeeping only prosperous with the use of honey bees?

10) A national flag is a flag which represents and symbolizes a country. The national flag is usually flown by the government of a country, but it can also be flown by its citizens. A national flag is designed with specific meanings for its colours and symbols. Historically, flags originated as military standards, which were used as signs on the battlefield. The practice of flying national flags outside of the context of warfare only became common in the early seventeenth century. A country's constitution will often describe the national flag. All national flags are rectangular, except for the flag of Nepal, which uses a unique triangular shape. The ratios of height to width vary among standard rectangular flags, but none is taller than it is wide. The flags of Switzerland and the Vatican City are the only national flags which are exact squares. The most popular colours used for national flags are red, white, green, and blue. Although the national flag is meant to be a unique symbol for the country it represents, many flags have very similar colours and designs.

1. Were flags originally made to show patriotism for one’s country?
2. Is orange among the most common flag colors?
3. Is it acceptable for a country to have a non-rectangular flag?
4. The form of the flag is often described in a nation’s law.

11) [TRANSLATE] The International Union for Conservation of Nature is an international organization working in the field of nature conservation and the sustainable use of natural resources. It is involved in data gathering and analysis, research, field projects, advocacy, and education. Its mission is to "influence, encourage and assist societies throughout the world to conserve nature and to ensure that any use of natural resources is equitable and ecologically sustainable". Over the past decades, the organization has widened its focus beyond conservation ecology and now incorporates issues related to sustainable development in its projects. Unlike many other international environmental organizations, it does not aim to

mobilize the public in support of nature conservation. Instead, the organization tries to influence the actions of governments, business and other stakeholders by providing information and advice, and through building partnerships. The organization is best known to the wider public for compiling and publishing the "Red List of Threatened Species", which assesses the conservation status of species worldwide. Today, the organization employs approximately one thousand full-time staff in more than fifty countries.

1. The International Union for Conservation of Nature is different from Green Peace.
2. This organization is mostly run by volunteers.
3. Its main goal is to raise public awareness on nature conservation.
4. This organization publishes the Red List of Threatened Species.

12) [TRANSLATE] A vehicle registration plate is a metal or plastic plate attached to a vehicle for official identification purposes. All countries require registration plates for road vehicles such as cars, trucks, and motorcycles. Whether they are required for other vehicles, such as bicycles, boats, or tractors, may vary by jurisdiction. The registration identifier is a series of letters and digits that uniquely identifies the vehicle owner within the issuing region's vehicle register. In some countries, the identifier is unique within the entire country, while in others it is unique within a state or province. France was the first country to introduce the registration plate, in the late nineteenth century. Early twentieth century plates varied in size and shape from one jurisdiction to the next, such that if a person moved, new holes would need to be drilled into the automobile to support the new plate. Standardization of plates came in the late fifties, when automobile manufacturers came to an agreement with governments and international organizations.

1. The first registration plate was introduced in France.
2. The size of plates was standardized before World War II.
3. Every national vehicle register includes both road vehicles and boats.
4. A vehicle registration identifier may be composed of letters, digits, and symbols.

Norwegian version

1) I eldgamle romanske religioner og myter regnes Janus som guden for begynnelsen og porter. Han har en dobbel karakter og avbildes ofte med to ansikt, fordi han kan se inn i fremtiden og fortiden. Janus styrte starten og slutten av konflikter, og derav fred og krig. Dører til tempelet hans var åpen under krigstider og lukket under fredstider. Som portenes gud ble han også sett på som en som kunne gå inn og ut av husenes dører. Janus symboliserte ofte endringer og overganger, som for eksempel fremgangen fra en tilstand til en annen, fra en visjon til en annen, og fra de unges vekst inn i voksen alder. Derfor var også Janus en gud som ble tilbedt i begynnelsen av høsting og plantetid, samt ved brylluper, dødsfall og andre begynnelser. Janus ble aldri tildelt en spesifikk prest, men den mektigste av alle prester utførte hans seremonier. Janus representerte middelveien mellom barbarisme og sivilisasjon, mellom landlige og urbane steder, og mellom unge og vokse. De gamle grekerne hadde ingen gud tilsvarende Janus, som romerne hevdet kun var deres.

1. Var Janus en gud av porter og passasjer?
2. Symboliserte Janus nattetid?
3. Var dørene til Janus tempel åpent i fredstid?
4. Var Janus en gresk gud?

2) Våpen med de likeverdige betegnelsene våpenskjold eller våpenmerke, er et kjennetegn som består av en sammensetting av farger (tinkturer) og figurer som vanligvis brukes i tilknytning til en skjoldformet innramming. Et skjoldmerke kan i tillegg ha én eller flere figurer på eller utenfor skjoldet, slik som kongekronen på Norges riksvåpen, eller kongekrone, våpenkappe og ordenskjede i det norske kongevåpenet. Våpen som eget fagområde kalles heraldikk. Våpen er innenfor heraldikken et kjennetegn som føres av en stat, by, en institusjon eller en slekt.

I gamle fyrstevåpen og i slektsvåpen er det mest vanlig at våpenets elementer består av skjold, hjelm, hjelmklede og hjelmtegn, uten andre tilleggsfigurer. De sentrale heraldiske normer, eller ledende synspunkter i middelalderen og i dag, er at figurer i våpen skal være generelle og abstrakte. Figurene skal ikke være for steds- og tidsbundne, de skal være egnet

til sterk stilisering og et våpen bør helst bare inneholde én eller så få figurer som mulig. Norden fikk nye våpen etter omkring 1600- tallet og bar ofte en utforming som brøt med middelalderens heraldiske praksis.

1. Er våpenet sammensatt etter heraldiske regler?
2. Kan skjoldmerke har bare et figurer?
3. Fiket norsk folk nye våpen etter 1400?
4. Representere heraldikk våpen et land og nasjonalitet?

3) I konkurranseidrett beskrives doping som bruk av forbudte prestasjonsfremmende medikamenter av idrettsutøvere. Begrepet doping brukes mye av organisasjoner som regulerer sportslige konkurranser. Bruk av medisiner for å prestere bedre anses i stor grad som uetisk, og er derfor forbudt av de fleste internasjonale idrettsorganisasjoner, inkludert Den Internasjonale Olympiske Komité. Dessuten ser man at idrettsutøvere som iverksetter eksplisitte tiltak for å unngå å bli oppdaget, forverrer det etiske overtrampet med åpenlyst juks og bedrag. Til tross for mange overskrifter i det siste, er ikke doping et nytt fenomen – det er faktisk like gammelt som idretten selv. Fra bruk av stoff i eldgamle hestevognløp til nyere kontroverser i baseball og sykling, har populære synspunkt blant idrettsutøvere variert mye gjennom årene. De siste tiårene har myndigheter og idrettsorganisasjoner forsøkt å strengt regulere bruken av narkotika i idretten. De viktigste årsakene til dette forbudet er helserisikoen ved prestasjonsfremmende medikamenter, likestilling av muligheter for idrettsutøvere og det positive eksempelet rusfri idrett setter for publikum. Antidopingmyndigheter har gjentatte ganger understreket at bruk av prestasjonsfremmende medikamenter er i strid med "idrettens ånd".

1. Kan doping gi alvorlig helseskade?
2. Bruker idrettsutøvere doping for å roe seg ned før konkurransen?
3. Er doping bare et problem i bestemte idretter?
4. Ble bruk av doping først observert i antikke brytekonkurranser?

4) Pungulv var det største, kjente kjøttetende pungdyret i moderne tid. Pungulv lignet en hund eller en hyene, men den hadde et lendeparti som klart skilte den fra canidene. Den er ofte kjent som den Tasmaniske tigeren, på grunn av den stripete korsryggen, eller den Tasmanske ulven (på grunn av dens hundeaktige utseende, trekk og egenskaper). Den kom fra det kontinentale Australia, Tasmania og New Guinea, og døde trolig ut på nittenhundretallet. Pungulven var ett av bare to pungdyr hvor begge kjønnene hadde pung. Hannen hadde en pung som beskyttet de ytre forplantningsorganene når han løp gjennom grove kratt. Pungulven blir beskrevet som et formidabelt rovdyr på grunn av sin evne til å overleve og jakte byttedyr i ekstremt tynt befolket områder. Pungulven var ekstremt sjelden eller utdødd på det australske fastlandet før den britiske bosetningen av kontinentet, men den overlevde på øya Tasmania. Intensiv jakt oppmuntret av dusører får vanligvis skylden for utryddelsen, men andre medvirkende faktorer kan ha vært sykdom, innføringen av hunder og menneskelig inngrep i dets habitat. Til tross for den offisielle klassifiseringen som utdødd, rapporteres observasjoner fortsatt, men ingen er endelig bevist.

1. Har mannlige pungulver pung?
2. Bidro introduksjon av sauer til utryddelsen av pungulven?
3. Blir pungulven sammenlignet med en tiger på grunn av halen?
4. Fantes pungulven i Asia?

5) Verdens miljødag ble etablert av FN i 1972, og markeres 5.juni hvert år for å øke bevisstheten omkring miljøspørsmål blant folket. Det er FNs miljøprogram som har ansvaret for markering av dagen, og dette gjør de i samarbeid med en ny storby eller land hvert år. Temaet som belyses henger sammen med miljøproblemer og utfordringer den aktuelle byen eller landet står overfor. Samtidig blir alle land, byer og lokalsamfunn oppfordret til å markere dagen der de er. I 2019 er hovedtemaet for verdens miljødag luftforurensing, og det er Kina som er vertskap for hovedmarkeringen. Fokuset skal rettes mot Asias storbyer, som har store problemer med luftforurensning. Verdens helseorganisasjon melder at det årlig dør 7 millioner mennesker som følger av luftforurensning, og tre millioner av disse bor i Asia. Luftforurensning og klimaendringer er to sider av samme sak, og sammen med blant annet utslipp fra fabrikker, fossilt brennstoff og gasser, bidrar også luftforurensning til global oppvarming.

1. Fokuserer verdens miljødag 2019 på eliminering av plastbruk?
2. Er klimaendringer og luftforurensing to sider av samme sak?
3. Feires Verdens Miljødag på ulike datoer hvert år?
4. Henger temaet som belyses på verdens miljødag sammen med landets miljøutfordringer?

6) En monokkel er en type korrigerende linse som før i tiden ble brukt for å kompensere for svakt syn. Monokkelen er en sirkulær linse som henger i en snor, og blir holdt på plass ved å knipe sammen øyegropen. Den andre enden av snoren kan festes til brukerens klær for å unngå å miste monokkelen. På slutten av det nittende århundre ble monokkelen som regel assosiert med velstående menn i overklassen. Kombinert med en lang frakk og en topphatt fullførte monokkelen bekledningen til en stereotypisk kapitalist på slutten av 1900-tallet. Monokler var også tilbehør til militære offiserer i denne perioden. Til tross for monokkelens popularitet på slutten av det nittende århundre, blir den sjeldent brukt i dag. Dette skyldes i stor grad fremskritt i optometri, som muliggjør bedre måling av brytningsfeil, slik at briller og kontaktlinser kan foreskrives med forskjellige og tilpassede styrker i hvert øye.

1. Brukte man hånden for å holde monokkelen på plass?
2. Ble monokler assosiert med et stereotypisk utseende?
3. Var stilendring grunnen til at man sluttet å ta i bruk monokler?
4. Ble monokler brukt av medlemmer av sosial og militær elite?

7) Vinsmaking er en sensorisk undersøkelse og evaluering av vin. Selv om utførelsen av vinsmaking er like eldgammel som produksjonen, er det sakte blitt etablert en mer formalisert metodikk fra senmiddelalderen og frem til i dag. Moderne, profesjonelle vinsmakere benytter en spesiell terminologi som er i stadig utvikling, og brukes til å beskrive spekteret av opplevde smaker, aromaer og generelle egenskaper ved en vin. De siste årene har det kommet frem resultater som utfordrer troverdigheten av vinsmaking for vineksperter og forbrukere.

Studier viser for eksempel at folk forventer at en dyrere vin har flere ønskelige egenskaper enn en billigere vin. Når vinsmakere får en vin som de blir fortalt er dyr, rapporterer de at den faktisk smaker bedre enn den samme vinen når de blir fortalt at den er billig. Andre studier viser at vinsmakernes vurdering kan påvirkes dersom de kjenner detaljer om vinen, for eksempel geografisk opprinnelse, omdømme eller annen informasjon. Objektiv vinsmaking forutsetter derfor at vinen blir servert blindt - det vil si uten at smakeren har sett etiketten eller flaskeformen. Blindtest kan også innebære å servere vinen i ett svart vinglass for å maskere vinens farge.

1. Er vinsmakere troverdige i vurdering av vinkvaliteten?
2. Er erfarne vinsmakere mindre fordomsfulle ovenfor vin enn en allmenn person?
3. Sikrer blindsmaking at vinsmaking blir mindre subjektivt?
4. Blir vin servert i et svart glass for å maskere fargen under vinsmaking?

8) Appelsinjuice er råsaft som oppstår når en appelsin presses tom for væske. I Norge, og i en rekke andre land, dominerer industrielt fremstilt juice. Blant disse skilles det mellom «ferskpresset juice» og juice laget fra konsentrat. Juice har en beskyttet varebetegnelse i Norge, og merkeforskriftene krever at næringsinnholdet i produkter som betegnes juice, er det samme som i den naturlige fruktsaften. Konsentrater kan tilsettes samme vannmengde som det som er fjernet, men ut over dette, tillates ingen tilsetningsstoffer i produkter som betegnes som juice. Dette inkluderer blant annet sukker og konserveringsmidler. Frukt- og bærprodukter med tilsetningsstoffer eller som ikke har fruktens naturlige og opprinnelige balanse mellom næringsstoffene betegnes som fruktnektar, limonade eller saft. Siden appelsiner inneholder store mengder vann, bruker mange produsenter i land uten egen appelsinproduksjon å fremstille appelsinjuice fra konsentrat. Etter innhøsting og pressing trekkes vannet ut av saften, og deretter fryses konsentratet. Konsentratet tar langt mindre plass, og er dermed mye billigere å frakte. I produksjonslandet blir konsentratet tint opp og deretter tilsatt like mye vann som ble fjernet før sending.

1. Er juice en beskyttet varebetegnelse i Norge?

2. Tillates det å tilsette sukker eller konserveringsmidler i produkter som betegnes som juice?
3. Er det vanlig at land uten egen appelsinproduksjon fremstiller appelsinjuice fra konsentrat og fryser det ned?
4. Blir vannet fjernet fra appelsinsaften før konsentratet sendes til produksjonsland?

9) Birøkt er stell av honningbier for å høste honningen de lager. De fleste som selger honning har det kun som biinntekt, men noen har det som hovednæring. Honningproduksjonen er svært avhengig av faktorer som temperatur og klima, og den norske produksjonen kan dermed variere fra rundt tusen til over to tusen tonn per år. Mens ville honningbier lever i trestammer og lignende, oppbevarer birøkterne sine bier i kuber. Kubene ble tidligere laget av halm eller tre, men nå lages for det meste av trykkfast isopor. Vokstavlene er det eneste inventaret i kuben. De er bygd opp av celler som vender vannrett ut fra en felles midtvegg. Byggematerialet er voks som biene produserer selv. Cellene er dels ammerom for yngelen og dels lagerrom for honning og pollen. Biene samler nektar i blomstene og lagrer den i cellene, hvor den omdannes til honning. Når honningen er moden, blir cellene forseglet med bivoks. På attenhundretallet vokste norsk birøkt kraftig. Dette skyldtes dels omveltningene i landbruket, som førte til en sterkere vilje til å prøve noe nytt, og dels nyvinninger i birøkten. Kubene ble sterkt forbedret ved at man fikk løse trerammer, der honningen kunne slynges ut ved at vokstavlene ble plassert i en sentrifuge. Dette førte til at man kunne hente ut honning uten å påføre stor skade på biene, og dermed kunne man beholde den viktigste råvaren – biene – fra år til år. Systemet med løse kasser er fortsatt nesten dominerende innenfor birøkt.

1. Honningproduksjonen er ikke avhengig av temperaturen.
2. Nå lages kubene for det meste av halm eller tre.
3. Biene produserer voks selv.
4. Nå kan man hente ut honning og beholde biene fra år til år.

10) Nasjonalflagg er et flagg som symboliserer et land. Et flagg er en tøyduk med bestemte farger, proporsjoner og eventuelt også symbolske figurer, som heises på stang. I nyere tid blir flagg i første rekke brukt som nasjonalmerke, men det brukes også som kjennemerke for byer, institusjoner, foreninger, firmaer og mer. Flagg har utviklet seg fra bannere og faner, som kan føre sin historie tilbake til oldtiden. For det enkelte lands nasjonalflagg skilles det gjerne mellom statsflagg, handelsflagg(handelsflåten) og orlogsflagg (marinen). Det nasjonale flagget har juridisk betydning og eventuelle misbruk vil ha rettslige konsekvenser. Stats- og folkerettslig har flagget særlig betydning for samferdselen til sjøs. Flagget er et tegn på hvor et skip hører hjemme, og hver stat fastsetter selv reglene for at et skip skal ha rett til å føre sitt flagg. Ifølge folkeretten må det imidlertid foreligge en reell tilknytning mellom flaggstaten og skipet. Både selvstendige staters flagg og flagg for nasjoner forstått som kulturelle fellesskap omtales som nasjonalflagg. I sistnevnte betydning kan man regne Catalonias, Englands eller Skottlands flagg, men også eksempelvis det samiske flagget.

1. Flagg brukes bare som nasjonalmerke.
2. Misbruk av det nasjonale flagget vil ha rettslige konsekvenser.
3. Hvert skip har rett til å føre sitt flagg.
4. Bare selvstendige staters flagg omtales som nasjonalflagg.

11) Den internasjonale foreningen for bevarelsen av naturen er en internasjonal organisasjon som har fokuset sitt på følgende områder: bevaring av naturen og bærekraftig utnyttelse av naturens ressurser. De deltar i innhenting og analysering av data, forskning, feltprosjekter, utdanning og endringsarbeid. Målet deres er å påvirke, samt oppmuntre og bistå samfunn rundt omkring i verden med å bevare naturen og sørge for at en eventuell utnyttelse foregår bærekraftig, økologisk og rettferdig. De siste tiårene har organisasjonen utvidet fokusområdet sitt utover det å kun bevare naturen økologisk. Den har nå utvidet fokus til å omhandle også bærekraftig utnyttelse av naturen. I motsetning til andre internasjonale naturorganisasjoner har ikke foreningen som mål å mobilisere det vanlige folket til å bistå i bevarelsen av naturen. Organisasjonen forsøker å påvirke regjeringene, næringsvirksomheter og andre relevante bedrifter ved å formidle informasjon og råd via alliansebygging. Organisasjonen er allmenn kjent for utgivelsen av «Listen med Utrydningstruede arter» som vurderer bevarelsen av

artene i hele verden. I dag har organisasjonen rundt ett tusen fulltidsstillinger i over femti land.

1. Foreningen for bevarelsen av naturen er ikke det samme som Green Peace.
2. Organisasjonen er hovedsakelig drevet av frivillige.
3. Hovedmålet er å fremme naturens bevarelse til allmenheten.
4. Organisasjonen utga listen med utrydningstruet arter.

12) Et kjennemerke for kjøretøy er et metall- eller plastikkskilt festet til et kjøretøy, som brukes til offisiell identifikasjon. Alle land krever kjennemerker for motorkjøretøy som biler, lastebiler og motorsykler. Om det kreves for andre kjøretøy som sykler, båter, eller traktorer kan variere i forskjellige rettsområder. Registreringsnummeret er en serie bokstaver og siffer som unikt identifiserer kjøretøyets eier i den utstedende regionens kjøretøyregister. I noen land er registreringsnummeret unikt for hele landet, mens det i andre land er unikt innenfor en stat eller provins. Frankrike var det første landet som introduserte kjennemerker, sent på attenhundretallet. Kjennemerker fra tidlig nittenhundretall varierte i størrelse og form fra ett rettsområde til et annet, slik at hvis en person flyttet, ville nye hull måtte bli drillet inn i bilen for å passe det nye skiltet. Standardiseringen av skilt kom sent på femtitallet, da bilprodusenter kom til enighet med myndigheter og internasjonale organisasjoner.

1. Det første kjennemerket ble introdusert i Frankrike.
2. Størrelsen på skiltene var standardisert før andre verdenskrig.
3. Alle nasjonale kjøretøyregister inkluderer både motorkjøretøy og båter.
4. Et registreringsnummer for kjøretøy kan være bygd opp av bokstaver, siffer og symboler.

Appendix 2: English reading comprehension test

1) Samuel Morse, best known today as the inventor of Morse Code and one of the inventors of the telegraph, was originally a prominent painter. While he was always interested in technology and studied electrical engineering in college, Morse went to Paris to learn from famous artists of his day and later painted many pictures that now hang in museums, including a portrait of former President John Adams. In 1825, Morse was in Washington, D.C., painting a portrait of the Marquis de Lafayette when a messenger arrived on horseback to tell him that his wife was gravely ill back at his home in Connecticut. The message had taken several days to reach him because of the distance. Morse rushed to his home as fast as he could, but his wife had already passed away by the time he arrived. Grief-stricken, he gave up painting and devoted the rest of his life to finding ways to transmit messages over long distances faster.

The main purpose of this passage is
A. to outline Morse's biography
B. to describe Morse's family life
C. to introduce a particular invention by Morse
D. to compare Morse's life in Paris and Washington, D.C.

Morse left the art world and helped to invent the telegraph because he
A. was tired of painting
B. wanted to communicate with people far away
C. experienced a personal tragedy in his life
D. was fascinated by science

2) Leonardo da Vinci is not only one of the most famous artists in history, but he was also a botanist, a writer, and an inventor. Even though most of his inventions were not actually built in his lifetime, many of today's modern machines can be traced back to some of his original designs. The parachute, the military tank, the bicycle, and even the airplane were foretold in

the imaginative drawings that can still be seen in the fragments of da Vinci's notebooks. Over five hundred years ago, this man conceived ideas that were far ahead of his time.

The author of this passage is praising da Vinci primarily for his
A. artistic talent
B. intelligence
C. foresight
D. fame

Among those listed in the passage, the common theme among da Vinci's designs is
A. architecture
B. transportation
C. optics
D. agriculture

3) The Amazon Rainforest is one of the most important ecosystems in the world. However, it is slowly being destroyed. Areas of the rainforest are being cleared for farms and roads, and much of the wood is also being harvested and sold. There are several compelling reasons to protect this area. First, a significant number of pharmaceuticals are made from plants that have been discovered in the rainforest, and it's quite possible there are still important plants that have not yet been discovered. Secondly, the rainforest provides a significant portion of the world's oxygen and also absorbs great amounts of carbon dioxide. Without rainforests, global warming could accelerate.

The main purpose of the passage is
A. to present the major reasons why the Amazon Rainforest is being destroyed.
B. to explain why the Amazon Rainforest should be protected.
C. to discuss how the rainforest has helped in the development of medications.
D. to argue that rainforest destruction is a major cause of global warming.

One contributing factor to the destruction of the rainforest is
A. construction of large dams
B. increase in tourism

C. logging
D. medicinal needs

4) Howard Gardner was a psychologist best known for developing the theory of multiple intelligences. Basically, the theory states that the idea of general intelligence or overall intelligence is somewhat inaccurate. This is because people often show intelligence in different areas. He argued that there are actually different types of intelligence. One type of intelligence that Gardner identified was interpersonal intelligence. People who possess this type of intelligence relate and interact well with others. Intrapersonal intelligence, on the other hand, implies that people are in touch with their own feelings. They enjoy thinking about theories and developing their own thoughts and ideas. People who have linguistic intelligence learn best by taking notes and reading textbooks. These people usually excel in traditional academic environments, as many academic subjects stress these types of activities. The other types of intelligence are kinesthetic, musical, spatial, and logical.

The main scientific contribution of Gardner is
A. forming an alternative for the theory of general intelligence
B. developing teaching methods for people with different types of intelligence
C. helping those who were previously considered intellectually disabled
D. highlighting the role of interpersonal intelligence

We can conclude from the passage that
A. Gardner believed that linguistic intelligence was the most desirable type to have.
B. most people who have a high level of intrapersonal intelligence do well in school.
C. people who have a high level of interpersonal intelligence work well in groups.
D. people who have mathematical intelligence would do the best on a standard IQ test.

5) The Internet has made life a whole lot easier for many people, but being online also brings with it very real risks. Hackers can steal personal and financial information. There are several precautions that computer users can take to minimize the level of risk that is involved with being online. One of the most obvious safety precautions is to purchase a good anti-virus and anti-spyware program. Passwords are also a very important part of online security, and

several tips can help users create more secure passwords. First, they should be something that can easily be remembered, but they should not be something others can guess easily. Your first or last name, phone number, or the name of your street are all bad choices, as people could learn this information quite easily. Longer passwords are more secure, and those that use a mixture of upper and lower case letters and a combination of letters and numbers are more secure than those that do not. Finally, passwords should be changed often. This can make remembering them more difficult, but the extra effort is worth the added security.

The main purpose of this passage is to
A. outline important considerations for passwords.
B. discuss the societal changes associated with Internet use.
C. talk about the importance of anti-virus programs.
D. discuss why certain types of passwords shouldn't be used.

According to the passage, changing passwords often is considered
A. beneficial, as it reduces chances of hacking
B. beneficial, as often change helps memorize passwords better
C. detrimental, as it may lead to often forgetting
D. detrimental, as it may lead to overly simplistic codes

6) Many people fail to realize just how crucial getting a good night's sleep actually is. It is usually suggested that adults get about seven hours of sleep every night, and younger children should get even more. Sleep has several benefits. First, it is believed to improve memory. This is one reason why it is always preferable to sleep the night before a test rather than stay up for the entire night to review the information. On a related note, sleep also improves concentration and mental alertness. Those who get sufficient sleep are able to concentrate on work tasks better and also react faster when they are driving a car, for example. Finally, people who get enough sleep have better immunity against illness. The reason for this is not fully understood, but researchers believe that an increase in the production of growth hormone and melatonin plays a role.

The main purpose of this passage is
A. to discuss how much sleep people should get.

B. to talk about the benefits of sleep.
C. to present strategies for improving memory and concentration.
D. to identify which hormones can boost immunity.

According to the passage, a large portion of the population
A. recognizes the benefits of sleep, but ignores them
B. cannot follow the recommendations regarding sleep times, due to financial reasons
C. prefers to sleep as much as possible
D. is unaware of the critical implications of sufficient sleep

7) A bird's feathers are extremely important, and when they clean and smooth them, it is known as preening. Birds in the wild preen their feathers on a regular basis. This is true of most captive birds as well, but not all. For example, some birds do not preen their feathers at all. This problem is most common in birds that are taken from their mothers at a very young age. Presumably, the absence of preening is due to the fact that they were never shown how to do it properly. A more common problem among captive birds is excessive preening. Some birds may pull out large numbers of their feathers or bite them down to the skin. It should be noted that wild birds never exhibit this kind of behavior. There are several suggestions about how the problem of excessive preening can be solved, such as giving birds baths or placing them in an area that has more activity to prevent boredom. However, these measures are often not sufficient to solve the problem.

The purpose of the passage is
A. to compare captive birds to wild birds.
B. to give an overview of abnormal preening in birds.
C. to discuss why preening is important.
D. to explain how excessive preening problems can be solved.

According to the passage, the most likely cause for the absence of preening is
A. being born in captivity.
B. being a large bird.
C. excessive bathing.
D. not getting an example from a parent.

8) Hibernation in animals is an extremely fascinating phenomenon, one that biologists are not yet close to understanding fully. However, it is quite easy to understand why animals hibernate during the cold winter months. Usually, it is because their food is quite scarce during this time. Animals that are herbivores will find the winters extremely tough, because all of the vegetation will have died off by the time winter arrives. Hibernation is essentially a way of dealing with this food shortage. Animals like birds rely on seeds and small insects for sustenance. Obviously, these will also be quite scarce in the winter when the ground becomes covered and frozen. Many birds address their upcoming food shortage in quite a different way: they migrate to warmer areas where their sources of food will be plentiful.

The main reason animals hibernate is
A. to avoid food shortages that occur during the winter months.
B. to avoid the harsh weather that occurs during the winter months.
C. to cut down on their food consumption during the winter months.
D. to save energy for the breeding season which typically occurs in the spring.

According to the passage, birds
A. often hibernate much like mammals
B. are less impacted by extreme weather conditions
C. have different ways of dealing with the winter conditions
D. change their food intake during winter

9) At one time, the use of leeches to treat medical problems was quite common. If a person suffered from a snake bite or a bee sting, leeches were believed to be capable of removing the poison from the body if they were placed on top of the wound. They have also been used for blood letting and to stop hemorrhages, although neither of these leech treatments would be considered acceptable by present-day physicians. Today, leeches are still used on a limited basis. Most often, leeches are used to drain blood from clogged veins. This results in little pain for the patient and also ensures the patient's blood will not clot while it is being drained.

The main purpose of the passage is
A. to explain how leeches can be used to remove poison from the body.
B. to compare which uses of leeches are effective and which are not.

C. to give an overview of how leeches have been used throughout history.
D. to discuss the benefits of using leeches to treat blocked veins.

In the past, leeches were often used as a way to
A. remove venom after animal bites
B. cure influenza and other viruses
C. control and reduce pain
D. perform exorcism by priests

10) When online file-sharing programs emerged, the music industry changed forever. Perhaps the first widely-used music file sharing program was Napster. It allowed users to sign up to use the service at no charge. Then, they could download music files from other users all over the world by simply typing in what song or album they wanted. Obviously, this was bad news for music artists and record labels because they were not making any profits from downloaded music. Eventually, Napster was shut down. While it later reinvented itself as a paying service, other free music-sharing sites cropped up almost immediately. Even though several sites and individual users have been charged, there are still countless individuals who log onto these sites to obtain free music.

The main problem associated with peer file-sharing sites is
A. they prevent artists and labels from earning money.
B. there are too many of them currently in existence.
C. it is hard to locate users and lay criminal charges against them.
D. they allow users to sign up for the service free of charge.

After Napster was shut down, peer file-sharing
A. became less available, since users did not know where to look for files.
B. became less common, since more users became wary of prosecution.
C. became more common, due to the publicity of such services.
D. was not dramatically affected, due to the emergence of similar services.

11) The pencil is a modern-day version of a centuries-old writing implement. Around 1560, an Italian couple designed the modern, wood-encased pencil. Their creation was flatter and more compact than the pencils we use today. Their plan involved hollowing out a stick of wood and inserting a stick of graphite into it. Shortly after, a better technique was discovered: two wooden halves were carved, a graphite stick was inserted, and then the halves were glued together, which is also how pencils are currently made. Although many people refer to the graphite inside pencils as "lead", they have always been made with graphite; however, the paint on the wood that surrounded the graphite was, at one time, lead-based.

According to the passage,
A. lead has only been used in pencils for a short while
B. today's pencil design is similar to that of the 16th century
C. today's pencils are made by scraping out sticks of wood
D. graphite is not a major component of pencils

The main purpose of the text is
A. to discuss the adverse effects of lead
B. to provide a history of the pencil
C. to describe the modern technology of pencil-making
D. to outline the biography of inventors of the pencil

12) Technology is rapidly expanding the scope of capabilities for both professional and personal use; such is the case with smart phones. Professionals now have mobile devices available to them capable of digital media, internet access, phone communication, multi-person scheduling and office tools for documents and presentations. Businesspeople that are often mobile may maximize the use of these critical features on smart phones. Individuals who simply enjoy the luxury of multi-function devices often use these devices for frivolous pursuits such as downloading catchy ring tones, instant messaging about the latest gossip and looking up the world record for most cans crushed on one's head during the Superbowl. This fusion of capabilities and increased availability of such devices could be a sign of a growing blend in society between work and personal life, or individuals could simply be taking a luxurious approach to their connectivity in personal lives.

What is the purpose of the conclusion sentence?

A. Draw a conclusion about the capabilities of smart phones

B. Assume where technology is headed and how it will affect society

C. Comment on human connectivity through the use of smart phones

D. Present two possible explanations for the growing popularity of smart phones

The passage does not shed light on

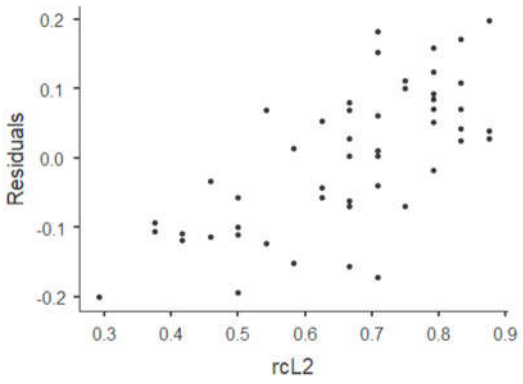
A. how technology changed communication patterns

B. how smart phones are used in business

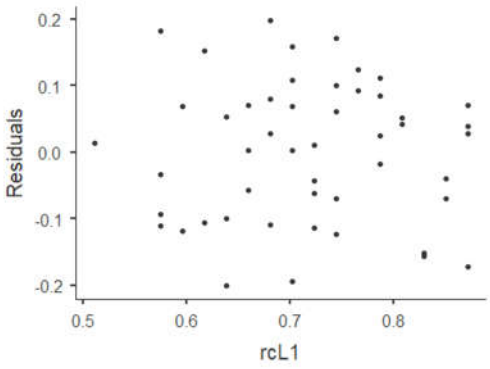
C. the use of smart phones for entertainment

D. the history of mobile phone devices

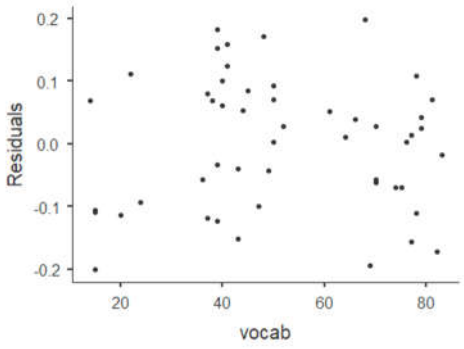
Appendix 3: Residuals plots



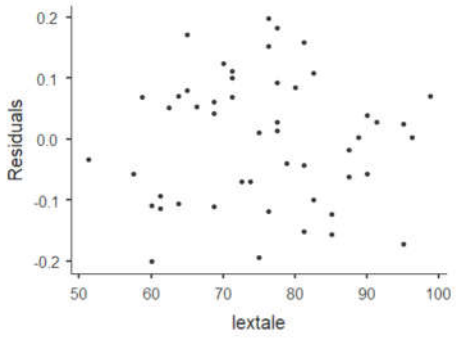
Residuals plot for L2 reading comprehension



Residuals plot for L1 reading comprehension



Residuals plot for vocabulary size test



Residuals plot for LexTale

Appendix 4: Consent Form

Vil du delta i forskningsprosjektet
«Eye movements in reading Norwegian and English»?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å forstå hvordan øynene beveger seg når vi leser på morsmål og fremmedspråk. I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Formålet med prosjektet er å studere forskjeller mellom morsmål og fremmedspråk når det gjelder øyebevegelser under lesing. Øynene beveger seg raskt fra ord til ord når vi leser, og fokuserer på nesten hvert ord så lenge som nødvendig for å gjenkjenne det. Stort sett ser det ut til at øyebevegelsene er like uansett språk, men det har ikke blitt systematisk forsket på hvilke språkegenskaper som avgjør hvor raskt og hvor stabilt øynene beveger seg. Studiet vårt er en del av et stort internasjonalt forsøk som samler data fra voksne som leser på flere språk med ulike skrivesystemer. Øyebevegelser under lesing sammenlignes i de ulike språkene og det undersøkes i hvilken grad de er påvirket av språk- og leseferdigheter.

Det er 3 masterstudenter som samarbeider for å samle inn og analysere de norske dataene og skal skrive masteroppgaver basert på dem.

Dataen blir samlet anonymt og skal offentliggjøres slik at forskere over hele verden skal kunne bruke data fra alle språk til videre forskning. Det betyr at din data kommer til å legges ut på nettet og vil være tilgjengelig uten tidsbegrensninger.

Hvem er ansvarlig for forskningsprosjektet?

Institutt for spesialpedagogikk ved Universitetet i Oslo er ansvarlig for den norske delen av prosjektet.

Det store internasjonale prosjektet styres av The Center for Advanced Research ved McMaster University i Canada.

Hvorfor får du spørsmål om å delta?

Data samles fra voksne lesere med norsk (Bokmål) som morsmål. Utvalget er trukket uformelt blant bekjente av studentene som jobber ved prosjektet.

Hva innebærer det for deg å delta?

Hvis du velger å delta i prosjektet, innebærer det at du fyller ut et spørreskjema og gjennomgår flere språk-, kognisjon- og lesetester. Det vil tilsammen ta deg ca. 1 time.

Spørreskjemaet inneholder spørsmål om språkvansker og engelsk språkkunnskaper, samt vokabular. Svarene dine fra spørreskjemaet blir registrert elektronisk.

Testene undersøker ferdighetene dine til å løse visuelle og verbale oppgaver, samt lese høyt raskt og stave på norsk og engelsk. Svarene dine på testene blir registrert elektronisk.

Prosjektet innebærer videre at du leser en rekke tekster på skjermen mens øyebevegelsene dine registreres. Det er 12 tekster på norsk og 12 på engelsk. Etter hver tekst skal du svare på noen få spørsmål angående tekstene. Denne oppgaven vil ta ca. 1 time til sammen (omtrent en halv time hvert språk).

Lesing og tester på morsmål (norsk) kommer først, og så kommer lesing og tester på engelsk etterpå.

Det er frivillig å delta
Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykke tilbake uten å oppgi noen grunn. Alle opplysninger om deg vil da bli anonymisert. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.
Dataen lagres helt anonymt, som gjør det umulig å finne tilbake til dine spesifikke resultater. Du har derfor ikke mulighet til å trekke deg etter at datainnsamlingen er gjennomført.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger
Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

- Navn og kontaktopplysninger blir ikke registrert utenfor dette samtykkedokumentet. I stedet for navn bruker vi en kode som på ingen måte er tilknyttet navnet.
- Anonymisert data kommer til å legges ut på en spesiell database utenfor EU der hvem som helst kan få tilgang til dataen uten tids- eller bruksbegrensninger. Dette gjelder både øyebevegelsesdata, spørreskjema og testresultat.
- Deltakerne vil ikke kunne gjenkjennes i databasen eller i publikasjoner.

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?

Prosjektet skal etter planen avsluttes høsten 2020. Prosjektdataen er helt anonyme og kommer til å være tilgjengelig på nettet i uavgrenset tid.

Dine rettigheter
Så lenge du kan identifiseres i datamaterialet, har du rett til:
- innsyn i hvilke personopplysninger som er registrert om deg,
- å få rettet personopplysninger om deg,
- få slettet personopplysninger om deg,
- få utlevert en kopi av dine personopplysninger (dataportabilitet), og
- å sende klage til personvernombudet eller Datatilsynet om behandlingen av dine personopplysninger.

Hva gir oss rett til å behandle personopplysninger om deg?
Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra Institutt for spesialpedagogikk ved Universitetet i Oslo har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

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

- Institutt for spesialpedagogikk ved Athanasios Protopapas, på epost (athanasios.protopapas@isp.uio.no) eller telefon: 22 85 77 05.
- Vårt personvernombud: Maren Magnus Voll, på e-post: personvernombud@uio.no
- NSD – Norsk senter for forskningsdata AS, på epost (personverntjenester@nsd.no) eller telefon: 55 58 21 17.

Med vennlig hilsen

Professor Athanasios Protopapas

Prosjektansvarlig og veileder

Areti Kalaitzi, Sara Fonseca, Veronica Tønnesen

Masterstudenter