The influence of English proficiency and native language reading on reading English

A quantitative study on the average looking time per word in English reading when considering Norwegian reading and English proficiency skills

Sara M. da Fonseca

Master of Philosophy in Special Needs Education

40 Credits

University of Oslo

Faculty of Educational Sciences

Department of Special Needs Education

June 30th, 2020

The influence of English proficiency and native language reading on reading English

A quantitative study on the average looking time per word in English reading when considering Norwegian reading and English proficiency skills

2020

The influence of English proficiency and native language reading on reading English

A quantitative study on the average looking time per word in English reading when considering Norwegian reading and English proficiency skills

Sara Maria da Fonseca

http://www.duo.uio.no/

Print: University of Oslo

ABSTRACT

The purpose of this study was to investigate the predictability of reading time in English as a second language (L2) in Norwegian young adults (N=50) between the age of 19-30. Native language (L1) looking time and proficiency scores from spelling and vocabulary knowledge were used to predict L2 looking time in milliseconds per word (mspw). The results of the study reported that Norwegian readers exhibited slower reading speed in L2 English than in their L1 (M=235 versus M=218). L1 Norwegian looking time contributed with 59.9% to the prediction of L2 English looking time per word. These results have shown that L2 looking time and L1 looking time are in line with previous findings on the influence of native language in second language reading. L2 proficiency scores in vocabulary and spelling contributed with additional 18.9% to the prediction model. Years of education was not significant to determine looking time in second language as it presented a common variance of .3%. The proficiency scores on spelling and vocabulary knowledge, L1 looking time and years of education, significantly predicted L2 looking time (explained variance, R^2 =.793).

.

ACKNOWLEDGEMENTS

First, I want to express my gratitude to Athanasios Protopapas for being a great supervisor. Thank you for always expressing his insight into this study, for all the questions he took the time to reply and the patience to explain to me what at times, seemed like an impossible thing to understand. A big thank you to Laoura Ziaka (Department of Special Needs Education, University of Oslo) for the time she took to help us.

I would also like to thank Kelly Nisbet (McMaster University, Canada), Anders Lunde (Department of Special Needs Education, University of Oslo) and Vasiliki Diamanti (Department of Special Needs Education, University of Oslo). Their help contributed for this project to run its course.

Also, thank you to my two colleagues that also took part in the cross-linguistic study. I am very grateful for the good teamwork between the three of us.

Lastly, I want to thank my closest friends, family and work colleagues for all the psychological support over the past two years.

Sara Fonseca

TABLE OF CONTENTS

List of Tables

List of Figures

Number of words: 19.958

1. INTRODUCTION

1.1. Background and Research Question

Reports from PISA (Programme for International Student Assessment) have identified that one in six Norwegian students leaves school unable to read adequately (OECD, 2013). A study on 319 teenagers, with 15-16 old, in Oslo, Norway, reported a prevalence of 10.4% in reading and writing disabilities (Green et al., 2009). Several researchers that investigated the impairment in reading and writing skills in Norwegian, agreed that reading difficulties are affected by skills involved in accurate and fluent word reading and spelling (Nergård-Nilssen & Hulme, 2014; Melby-Lervåg & Lervåg, 2013). With basis on those studies, if reading ability in first language (L1) depends on such linguistic skills, it can be argued if such effect also occurs in second language (L2). Considering that reading in native language (L1) is a complex process, the involvement of second language (L2) is not an easier task. The ability to read efficiently in L2, particularly in English, for academic purposes is considered a critical skill in many countries.

In L2 reading research, fluency has not yet received as much attention as it has in the field of L1 reading. According to Yamashita and Ichikama (2010), the lack of studies in L2 reading fluency is possibly because L2 researchers tend to focus more on reading comprehension rather than rate and smoothness of reading due to higher difficulty in L2 reading comprehension when compared to L1. Nonetheless, understanding fluency in second language is no less important, because L2 learners should be able to increase reading speed and still be able to comprehend L2 texts.

A fluent reader is characterized by the speed s/he can read texts while understanding the meaning of words (Rasinski et al., 2011). Literature supports that complex languages that require more words to transmit a message, have a higher reading rate, which is expressed in words per minute (wpm), whereas languages involving fewer words have a lower reading rate (Brysbaert, 2019). It considers the looking time readers need to comprehend text. Defining a rate that expresses fluency in reading is relevant as it can measure the time and effort required across tasks, which can be of great help to assess reading disabilities. Moreover, “reading rate is also important for psychological theories about the reading processes, individual differences in information processing, language

differences and task effects" (Brysbaert, 2019, p.44). That is, the amount of time needed to read is reflected by someone's knowledge of the language.

To measure proficiency however, based on the reader's individual differences, is a more complex process. In this thesis, we will focus on lexical knowledge skills, particularly vocabulary and spelling, to consider proficiency in L2 English. Studies that account the development of reading skills in native language (L1) and second language (L2) in Norwegian secondary school students (16-19 years old) showed that proficiency tends to be very high for this age group (Ibsen, 2002). Besides, results from PISA (Programme for International Student Assessment) have shown that L1 and L2 reading skills increased significantly over the years (OECD, 2013; Ibsen, 2002). A similar study showed that Norway scored second in a European reading assessment with L2 English (Ibsen, 2002). The results from the study show that Norwegians are overall skilled readers of L2 English.

Literature supports that acquisition of second language is strongly affected by one's native language (Verhoeven et al., 2019). Hypothetically, similarities in Norwegian and English may influence how well you can speak a second language. Nonetheless, it may be the case that one's ability to read in L2 does not match the same level of mastery in spoken language, despite being a fluent speaker. May that be the case for Norwegians? This study aims to shed some light on these issues by investigating if English looking time per word, as a second language of native speakers of Norwegian, can be predicted by the reader's levels of proficiency in vocabulary and spelling in L2. Furthermore, it is in the interest of this paper to analyze the correlation in L1 and L2 looking time in reading. Finally, it is intended to explore the predictability of reading looking time in L2 English when considering years of education, L1 Norwegian reading looking time, vocabulary and spelling proficiency scores. With that in mind, the following research question was formulated:

> *Can Norwegian looking time per word and English proficiency scores in vocabulary and spelling predict reading looking time in English?*

1.2. Description of terms

The abbreviations L1 (native or first language) and L2 (second language) are mentioned several times in the theoretical chapter as they consider previous research. It is important to note that only from the methodology section (Chapter 3), L1 and L2 exclusively refer to Norwegian and English, accordingly.

The variables English mean viewing time and Norwegian mean viewing time are often referred to as “looking time” and they refer to the same variable. Mean or average viewing time per word expresses the amount of time (in milliseconds) that a reader spends in one word. This means that if the reader spends longer time (larger number in milliseconds) in each word, the slower the reader. Similarly, if the reader needs less time to look at words it means that s/he is reading faster.

1.3. Structure

This thesis is divided in six chapters. The present chapter (Chapter 1) presents the background and topic of choice with article-based research that supports the pertinence of the study, followed by the research question. Chapter 2 starts by presenting the relevance of reading research in special needs education. Thereafter, theories on reading, reading fluency and reading rate are considered. Following, average looking time in first- and second language are defined along with the elements that influence the dependent variable of the study. The theoretical chapter concludes with a detailed description of eye tracking research in reading.

Research methodology is considered in Chapter 3. The chapter describes the design of the study, eye-tracking equipment, sample and data collection. Then, the variables used in the study are presented as well as the reliability and validity used to assess the measurements. After, the statistical methodology of this thesis is presented, followed by research ethics considerations and scientific principles. In Chapter 4, the results achieved from the descriptive analysis, the bivariate correlation analysis and the hierarchical multiple regression analyses are presented.

Chapter 5 discusses the validity and reliability of the measures used to collect data and the results of the study are discussed in light of relevant literature and scientific research findings. Lastly, conclusions of the thesis are drawn and suggestions for further research are presented in Chapter 6.

2. THEORETICAL BACKGROUND

2.1. Theories in Reading

2.1.1 Relevance of reading research in the Special Needs Education field

The ability to use language is a fundamental skill in everyday life. If we think about it, there are not many things that we would do without it. Reading research is especially important in the educational field because the use of language in written form has been used in schools for several decades (Fredrickson & Cline, 2015).

Reading abilities develop through practice, in other words the more one practices, the better reader s/he will be (Lyster, 2001). This statement is true even for people that struggle with reading which is why impairments in learning and literacy directly affect reading development. For example, learning difficulties related to meta-cognitive processes may affect working memory which may enable the learner to retain information and relate it to following knowledge (Fredrickson & Cline, 2015). In other words, the student would not be able to make sense of language over time as it requires systematic learning. A similar process occurs when learning a second language (L2), as new linguistic knowledge builds over one's native language (Cummins, 1979). Therefore, if struggling in native language, problems with second language are likely to occur. Also, literacy disabilities in reading progress. If a pupil lacks knowledge in language comprehension and word recognition (Gough and Turner, 1986) then the child will never be able to improve reading skills unless special strategies are provided (Fredrickson & Cline, 2015). Such effect also occurs in reading disabilities, such as dyslexia.

Lyster (2001) mentions that English is an irregular language when compared to Norwegian. That is because in English, the same string of letters can be read in different ways and similarly the same word can have different meanings (e.g. the two different meanings and pronunciations of the word *live*) while in Norwegian such event rarely occurs. Based on Lyster's (2001) considerations one may question if English can be considered a more difficult language to learn. If so, in learners that have English as an additional language (EAL) or as a second language (L2), the process of learning a new language may be even more complex (Fredrickson & Cline, 2015), especially if the student has reading disabilities.

There is not enough information about how exactly Norwegian language operates on the English reading. Studies in L2 English and L1 Norwegian reading may help understand the linguistic transfer between these two languages and consequently help students with special needs education work on strategies that improve reading in L2.

2.1.2. The Simple View of Reading

Reading is defined as the processing of textual information while retrieving the intended meaning of each word (Rayner et al., 2016).

The Simple View of Reading (Gough & Turner, 1986) expresses how reading is the product of decoding and comprehension (D X C = R). One cannot function without the other in order to acquire reading skills. Decoding refers to word recognition processes while comprehension pronounces language comprehension processes. The National Institute for Literacy (2008) defines decoding as the process of accurately and fluently translating written words into spoken language. Gough & Tunmer (1986) define language comprehension as the ability to give semantic meaning to spoken words. In order to measure language comprehension, assessment in vocabulary and word definition should be considered.

2.1.3. Reading fluency

In research about reading development, one of the most important variables to consider is reading fluency. Breznitz (2006) defends three different positions that define reading fluency. The first one approaches fluency as a measurable outcome of oral reading pre-skills. In other words, how fast a person can decode and connect words to establish a meaningful text. The second domain considers reading fluency as a linguistic and developmental outcome of word decoding and recognition, syntactic processing, semantic information and meanings. The last domain presents reading fluency as an effective systemic processing outcome that relays on biological and cognitive abilities. In fact, it can be considered that all domains interact with the position of fluency in reading development.

Similarly, Fraser (2007) considered that reading fluency can be achieved through the automatization of text-based reading components, specifically word recognition and linguistic skills.

Another interesting definition to consider is given by Rasinski et al. (2011), who described fluency as a “characteristic of reading that occurs when reader’s cognitive and

linguistic systems are developed to the extent that they can read with sufficient accuracy and rate to allow for understanding the texts" (p.287). In other words, fluency is considered a critical aspect of reading ability because it determines the readers' capacity to construct meaning.

Linguistic skills, such as vocabulary and spelling skills, are considered to be essential for language comprehension and decoding processes (Breznitz, 2006). Fluency in reading benefits from vocabulary because the larger the lexical knowledge, the easier it is expected to recognize words and attribute meaning to them. Through spelling, the reader can perceive individual letters and the correct orthography of words, facilitating word recognition fluency (Perfetti, 1985 as cited in Breznitz, 2006). Furthermore, these language skills have a significant role in reading fluency because if a reader is familiar with vocabulary, syntax and phrasing s/he is more likely to anticipate upcoming text in a sentence (Rasinski et al., 2011).

In fluency measures the text is read aloud only words that are pronounced correctly are counted (i.e., incorrect words are not counted). For combination of persons and passages that entail no (or every few) reading errors, the rate of reading aloud equals the fluency wpm (words per minute) metric, whereas for passages that are challenging to the readers and result in reading errors, fluency wpm is less than plain rate. However, for silent reading this distinction cannot be made, and only reading rate can be measured. There, it is important to use passages at the appropriate level for the metric to be meaningful.

2.1.3.1. Reading rate

Reading rate is seen as the outcome of reading skills (Breznitz, 2006). That is, reading rate is defined as a result of good decoding and comprehension ability in reading. Basically, faster reading rates reflect better decoding and comprehension skills while slower ones reflect the opposite (Carver, 1997, as cited in Breznitz, 2006).

Reading speed is very often expressed in research in wpm. There are several ways to calculate wpm as a rate in reading. The most common one is calculated by summing the number of words in a passage and divide them by the time spent reading the passage, in minutes (Rayner et al., 2016). In reading speed research, a very often discussed topic is how fast someone can normally read silently and what factors can influence speed reading (Brysbaert, 2019).

A study among English native speakers showed that, when reading easy materials for general comprehension, reading rates are between 280 to 300 wpm (Carver, 1990, as cited in Brysbaert, 2019). The scholar also noted that the reading rate decreased to 200 wpm when the passage is beyond the reader's easiest reading level or when comprehension would be assessed. In fact, it is self-explanatory that harder texts or being measured for understanding after reading, would increase the reader's attention to the passages which should reflect lower reading rates.

However, a recent meta-analysis on 132 studies of reading rate, estimated that in English, the average silent reading rate for adults is 238wpm for non-fiction passages with a range of 175 to 300 wpm, whereas for fiction passages the average reading rate was, with a range of 200 to 320 wpm (Brysbaert, 2019). After analyzing the results from 132 studies, Brysbaert noticed that for silent reading in English adult readers fall in a range of 175-300 wpm when reading fiction.

A study by Haynes and Carr (1990) investigated the relation between writing system knowledge with different aspects of reading skill, such as comprehension, speed, and the ability to learn new vocabulary from context, in native readers of Chinese with L2 English. The scholars found that slower L2 rates tend to be correlated with native languages (L1) that are more distant from English in language writing systems, than L1s that are closer to English.

Another study by Oller and Tullius (1973) investigated the reading speed in 50 foreign students with L2 English. In this sample, 21 different L1 languages were divided in two groups according to the language family. More specifically, one group to languages were L1 belonged to the Indo-European languages (IE) those whose L1 did not (NIE). The results shown that the IE group had a mean English reading rate of 227 wpm, the NIE group had a rate of 182 wpm. English is an Indo-European language which may explain why reading rate for the IE group is higher than the NIE group.

The important note to take from these studies, is that reading speed in L2 is determined by the level of knowledge in L2 (faster reading is associated with higher proficiency), the writing system and construct of the native language.

2.2. Average Looking Time in Silent Reading

The variable mean total looking time is expressed in milliseconds per word (mspw). It is a result of summing the time of all the viewings and then dividing that sum by the number of words in a passage. Furthermore, it can be considered a good measurement

to assess reading fluency because it is presumably a good index of the total effort associated with each word. This understanding is built on Just and Carpenter (1980), as cited in Štajner et al. (2017), thinking that when a reader looks at something, the information retrieved from it is being cognitively processed. Therefore, the amount of time spent while processing words should be considered when assessing reading. In fact, the importance of time associated with reading has been considered in eye-tracking research. Research with eye tracking measures has shown that time spent reading has been used to measure word properties in reading and can help detect reading disorders, such as dyslexia (Holmqvist et al., 2011).

However, it is important to note that the variable average looking time in silent reading is not considered a reading rate. As previously stated, reading rate considers speed in reading (Breznitz, 2006), that is the greater the number of words you read per minute, the faster the reader. On the other hand, higher numbers of viewing time represent longer time spent in each word which signifies as slower reader. The smaller the average looking time, less effort is associated with reading which suggests a faster reader. Average looking time values should range according to individual knowledge and proficiency in language.

2.2.1. L1 Looking time

L1 word viewing time consider the effort associates with each word when reading in native language. The native language of the sample used for this study was Norwegian (Bokmål). Norwegian Bokmål is the language used by 85% to 90% of the population in Norway (Vikør, 2005).

Readers are expected to read faster, that is having higher values of reading rate, in their native language (Brysbaert, 2019). However, a native reader should present smaller values on average looking time in their L1, which would acknowledge less time needed to identify and decode each individual word.

2.2.2. L2 Looking time

L2 looking time is defined by summing the time of all viewings and then divide them by the number of words in the English passages.

Reading speed in second language, L2, is slower when compared to native language, L1 (Brysbaert, 2019). In fact, a study on undergraduate thirty-three students, nineteen Dutch (L1) and English (L2) bilinguals and fourteen English native readers, compared

eye-movements patterns of bilingual readers in L1 and L2 (Cop et al., 2015). The results of the study have shown that reading rate was slower in L2 by 17%, when compared to the native language. A similar study between French (L1) and English (L2) speakers found that L2 reading rates were 30% to 40% slower than in L1 (Favreau et al., 1980). Considering that average looking time expresses the effort associated with each word, lower values of L2 looking time can be associated with better readers. Moreover, smaller distance between values of L2 and L1 looking time can signify fluency in both languages. However, based on the results from previous studies on reading rate, the values of looking time per word are expected to be higher for L2 as they should express lower fluency when compared with the native language, L1.

In order to study the predictability of L2 looking time, there are four factors that will be considered to examine the fluency in English as second language. The first factor relates to the writing systems of both languages used in this thesis. The second factor considers the linguistic transfer between L1 and L2 followed by the exposure to L2 English in the Norwegian schools. The last factor to investigate is how proficiency in L2 can influence reading fluency.

2.2.2.1. Writing systems

Most languages, such as English and Norwegian, have alphabetic written system. Attaining the understanding of the alphabetic code is critical to learn how to read. In alphabetic writing systems, the sound of letters is called phonemes. Phonemes slightly change between languages to approximately represent the sound of letter according to the language alphabet (Rayner et al., 2016).

Norwegian and English are part of the Indo-European family, within the Germanic family branch (Harbert, 2006). These two languages have the same reading direction (left to right) and share the Latin writing system, which means they share the same letters of the alphabet with disparities in letter pronunciation, according to the Encyclopædia Britanica (Editors of Encyclopædia Britanica, 2020)

Norwegian is a North Germanic language that has three more letters (æ, ø, å and they are pronounced like the vowels in bad, burn and ball, accordingly) than the Latin alphabet.

Alike so many countries, Norwegians has many dialects. However, in the present day they consider two “versions” of their language: Bokmål and Nynorsk, which directly translate into “book language” and “new Norwegian”. Norwegian Bokmål is the main

language in almost every school and is the predominant dialect in the area around Oslo and the eastern Norwegian lowland. Nynorsk is often used in oral speech in the interior and along the west coast of the country (Enander et al., 2020)

English is a West Germanic language with a range of dialects and linguistic varieties across pronunciation, morphology, vocabulary, orthography and grammar. There are several dialects of the language in spoken and written form, in European schools we are taught British English (Potter & Crystal, 2019).

English vocabulary is approximately a quarter Germanic (Old English, Scandinavian, Dutch, German) and several nouns and verbs are identical whether they came from Old English or Scandinavian (Potter & Crystal, 2019). English is not a mutually comprehensible with other Germanic languages due to differences in syntax, phonology and vocabulary (to a greater extent). There are however a few exceptions, where some languages (i.e., Dutch) show similarities with English in early stages of linguistic learning (Harbert, 2007). Nevertheless, the fact that Norwegian and English belong to the Germanic family, may justify the proximity of writing systems among the two languages. Due to this proximity, Norwegian as an L1 may explain some of the variance in L2 English reading (Brevik, et al., 2016).

The idea that similarities in writing systems may influence average looking time is not excluded because sharing the same alphabetic principle should facilitate linguist transfer between two languages (Verhoeven et al., 2019). An example of such event is sharing the same alphabet, which reflects similarities in orthography, that is "the accepted way of spelling and writing words" (McIntosh, 2013). In fact, even if the languages differ in their syntactical, phonological and morphological structure, singular effects on linguistic transfer in orthography can still be predicted on L2 English, with basis on L1 writing system (Haddad, 2019). Haddad's study (2019) considered two languages with different writing systems that shown evidence of influence in L2 with basis on L1 linguistic characteristics, specifically phonological skills. The study considered L1 as Hebrew and English as L2, with monolingual and bilingual participants in the sample. Haddad (2019) observed significant differences in phonological skills between the two groups and that the phonological distance between L1 and L2 affected phonological awareness in L2. This difference was more noticeable in tasks where participants were asked to identify phonological units that do not exist in their L1 phonological system.

2.2.2.2. Linguistic transfer from L1 to L2

The transfer in L1 and L2 samples is expected between alphabetic languages that are close in origin (e.g., Indo-European languages), like those used in this study, because they share a constituent knowledge in learning to read – the alphabetic principle (Melby-Lervåg and Lervåg, 2011). In other words, the writing system can facilitate the transfer between L1-L2.

The linguistic transfer between languages has been immensely studied in oral reading, reading comprehension or to investigate the effect of specific reading skills. However, research studies that consider the effect of L1-L2 transfer in silent reading rate are limited.

Olsen (1999) studied the results of L2 English writing in Norwegian (L1) learners in different elementary district schools in Norway. The results were a clear indicator that native language played an important role in L2 development. More specifically, students used words that exist in Norwegian and transformed them into English (code-switching) and transferred spelling, grammatical and syntactic rules by slightly adapting L1 Norwegian rules to produce a language that was neither Norwegian nor English, but rather a mixture of both.

Brevik et al. (2016) were the first researchers to do a systematic research that compares reading in Norwegian L1 and English L2. In this study, the scholars used the results of two national reading exams, in Norwegian and English, from the year 2010. The study considered the results of 10,331 students in 11th grade that attended public schools (87 in total). Norwegian L1 was assessed on a paper-based test while English L2 was a digital test. Both reading tests had the same construct and assessed language (vocabulary and grammar) and reading comprehension. Findings from this study showed that L2 English and L1 Norwegian have a close relationship and shared characteristics of reading in the L1 and L2. Brevik et al. (2016) found that L1 reading comprehension and proficiency was the strongest predictor of L2 reading proficiency. Hypothetically, this conclusion might be a result of the linguistic similarities between Norwegian and English. Nonetheless, readers are only expected to be equally efficient in reading in L2 and L1 when they have the same amount of exposure to L2 words as L1 native speakers have to L1 words (Brysbaert, 2019). Lastly, the regression models from Brevik et al. (2016) confirmed a strong positive relationship of up to 41% between L1 and L2 reading proficiency, and that L1 reading language and comprehension, age,

gender and different study programs in upper secondary school can account for 43% variance in L2 reading proficiency.

There are two important messages to take from this section. The first one relates to the linguistic transfer across L1 and L2 is facilitated by the writing system of the languages being studied. The second one is that reading in L2 requires skills that are developed in L1, and that the transference from L1 is noticeable in L2 reading. The shorter the distance between L2-L1 the easier it is to transfer native language skills into second language. Similarly, the larger the distance the harder it is to learn another language (Schepens, van der Slik, van Hout, 2013). As an example, if considering Norwegian and Swedish (two languages very close linguistically), it is easy to consider that a native Norwegian can more easily assess reading in Swedish then a native Norwegian to access Hebrew, a language that has a totally different writing system and reading direction.

2.2.2.3. Level of Education

In Norway, the age of acquisition of a language is directly related to the level of education as English is a compulsory subject from 1st until 10th grade (6 to 16 years old). Upper secondary school is optional and consists of three more years (11th to 13th). For students continuing to upper secondary school, English is a compulsory subject taught also in 11th grade, for vocational courses until 12th grade, and can be thought as an elective subject (Norwegian Ministry of Education and Research or Kunnskapsdepartementet [KD], 2013). Thereafter, if students proceed to higher education, the exposure to English derive from published research and literature which requires a higher level of proficiency in reading.

Norwegians that have completed the minimum education are expected to have received 588 hours of training in English. By the time Norwegians finish school, at the age of 16, they should be able to read fluently, create meaning from different types of texts, despite of length and complexity, executing and processing reading English-language texts for different purposes and to apply reading strategies adapted to the purpose of reading in increasingly demanding texts (The Norwegian Directorate for Education and Training or Utdanningsdirektoratet [UDIR], 2013).

As previously stated, decoding is necessary for reading comprehension and the developmental skills in second language reading. The age of readers influence decoding experience (Jeon & Yamashita, 2014). In that sense, the level of education, usually

associated with the readers age, relates to L2 reading comprehension. Reading comprehension in L2 is also associated with the growth of skills through experience which means that the higher the level of the education, the bigger is meant to be the experience and knowledge in reading (Breznitz, 2006).

A study by Brevik et al. (2016) carried out in Norwegian Upper Secondary schools showed that students in general studies have better L2 reading comprehension scores than students from vocational courses. It was previously mentioned that the amount of time Norwegians spend learning English by the time they finish school should produce a fluent reader capable of understanding text at the same level. The variation of the results poses a threat to the reliability of the measures or can assume that students of general education, and that intend to proceed to university, have a better understanding of L2 English.

Studies by Verhoeven et al. (2019) on neurocognitive processes, indicate that L2 reading processes towards fluency depend also on the age of acquisition of L2. This is based on the fact that the cognitive processes associated with reading develop in early ages facilitate language.

Van den Bos, Zijlstra, and Spelberg (2002) addressed developmental relations between naming and word reading speed to determine if these developed at various age levels. Naming speed is important for reading fluency as it undertakes the reader's ability to quickly retrieve the words. Results from the study confirmed that naming speed increases as a function of age. In other words, faster recognition of words is related to increasing in reading speed which simultaneously, increases reading rate. However, it is important to note that such abilities, after a certain age, tend to decrease with time (Brysbaert, 2019).

Overall, years of education are associated with the development of cognitive skills, word recognition and reading comprehension according to the age of the readers. These can in their means affect average looking time, because if the readers do not have enough experience and knowledge to process the words, more time will be spent in each of them, increasing viewing time. Basically, readers that have completed at least 10 years of education in Norway are supposed to understand L2 texts to an extent that does not affect reading fluency.

2.2.3.4 Proficiency in L2

It is important to understand how proficiency in L2 may affect reading fluency and why it is of relevance to be measured in order to predict reading looking time in L2.

The *Cambridge Advanced Learner's Dictionary* (McIntosh, 2013) defines proficiency as "the fact of having the skill and experience for doing something". Thus, proficiency in L2 defines the experience and skills in second language.

There are several ways to measure proficiency in L2. All measurements focus on the learner's understanding of second language in several components: lexical decision, word knowledge, orthographic knowledge, syntax knowledge, grammar knowledge, listening, reading and writing comprehension, among many others (Gutierrez, 2016). You need to be highly proficient in a language in order to be able to understand it and interpret the meaning of each work, in written or oral form. At the same time, the better someone is at a language, that is the better vocabulary, syntax, grammar and general understanding one has of the specific language, the better the reader. In a simpler way, the level of proficiency should reflect better readers. So, better proficiency in a language should also relate to how fast you would be able to read. The two are normally associated because becoming increasingly proficient typically includes improving reading skills as well.

The results from a cross-linguistic study (Fraser, 2007) showed that L2 proficiency can predict one's performance in reading tasks that involve memorizing the content of the texts.

In other words, better levels of proficiency showed better understanding of the passages while slower reading rates of L2 learners supported lower results comprehension tasks. Fraser used a listening test to assess knowledge in vocabulary, morphology, syntax and discourse in L2 English. The reason to use an oral format tests to assess proficiency was because previous knowledge showed that reading comprehension is a product of mental representations of the sounds of words that the reader creates while s/he reads (Harris, 1986).

Proficiency in L2 in reading can also have a great impact in L2 reading comprehension. A study on French (L1) learners of English (L2), examined two different levels of proficiency in L2, a lower-intermediate group and an upper-intermediate group (Walter, 2004). The results showed that the lower proficiency group showed difficulty in making sense of some of the texts in L2 but not in their L1, which affected comprehension in

L2. It should not be surprising that lower skills in L2 related negatively with L2 reading.

In short, this section means to acknowledge that slower L2 reading rates have been associated with lower levels of proficiency in L2 as the reader needs longer time to process text.

2.3. Vocabulary and spelling

According to *Cambridge Advanced Learner's Dictionary*, vocabulary can be defined as "all the words which exist in a particular language or subject" or "all the words known and used by a particular person" (McIntosh, 2013). The first definition refers to the word level which includes everything from one particular word to all the words in a language, while the second definition comprises the reader's *lexicon,* that is the word knowledge of the reader. For L2 English readers the preferred way to define vocabulary is by considering words that can be translated with reference to the reader's native language (Nation, 1990). In other words, how does one translate an English (L2) word into Norwegian (L1).

Familiarity of words or having vocabulary knowledge can have several definitions, ranging from simple recognition of a word to a complete mastery of a word. To master vocabulary, knowledge of word meaning, orthographical and phonological form, collocations, associations, grammatical behaviors, register and frequency are necessary (Nation as referred to in Schmitt, 2001). Various studies have attempted to estimate the actual number of words L2 learners need to know to comprehend text. Scholars have theorized adult L2 learners needed the same number of words in their lexicon as adult native speakers (Goulder, Nation, & Read,1990, as cited in Guo & Roehrig, 2011). According to Laufer (1997), as cited in Guo and Roehrig (2011), about 3,000-word families or 5,000 individual word forms are necessary for L2 learners' minimum comprehension. Nevertheless, is important to note that absolute word knowledge is an unrealistic goal for most native readers, even more L2 readers.

Spelling is defined by the ability to recognize or reproduce a correct sequence of letters in oral or written form (Perfetti, 1997). The process of spelling implicates the integration of phonological and alphabetic skills of beginning reading (Santoro et al., 2006).

Spelling reflects the general principles of the writing system and its design, and the specific orthography that embodies the writing system and its distinctive features.

Spelling instruction strengthens alphabetic understanding by emphasizing the names of letters, the sounds they most frequently symbolize, and how to group letters to form graphemes (Santoro et al., 2006). More specifically, spelling is a linguistic skill that integrates and relies on several layers of knowledge, such as phonological awareness of speech-sounds in words, morphological awareness, semantic knowledge, and orthographic knowledge of the letter sequences and patterns that are used to spell words (Moats, 1984).

Orthography in English is quite irregular. In spelling these irregularity makes it complicated to predict which letter to use to orally spell a sound (Kohnen et al., 2009). In other words, in English there are many letters that change sound and some sounds are actually not represented by its usual letter(s). Another important source of irregularity in the English is the attributed meaning in similar words that have related meanings and spellings despite sounding quite different, such as the example in the words *please, pleasant, pleasure,* where the '*ea'* represents two different sounds (Templeton et al., 2007, as cited in Kohnen et al., 2009).

2.3.1. Vocabulary knowledge and spelling to assess language proficiency in L2

In the educational field, a preference has been given to the use of standardized tests because they guarantee equal conditions of administration, scoring procedures and interpretations. Furthermore, standardized tests were constructed to increase the reliability of the scores by maintaining constant as many factors as possible (Olson & Sabers, 2008).

Language proficiency in L2 can be assessed by measuring lexical precision in vocabulary and spelling (Andrews & Hersch, 2010). Lexical precision in L2 English, measures word recognition by assuming that the reader is able to understand the representation of sounds for every word and their meanings, and for pseudo words (nonwords) the reader is supposed to breakdown how the string would sound despite the fact that is not a real word.

Knowledge of spoken vocabulary is an important prerequisite to acquire English as a second language (e.g. Lesaux & Siegel, 2003). The Lexical Test for Advanced Learners of English (LexTALE) is a nonspeeded English lexical decision task and is meant to discriminate between highly proficient speakers. LexTALE is meant to evaluate vocabulary knowledge. Even so, the scores of the test were validated to measure general English proficiency (Lemhöfer & Broersma, 2012). The use of this test in a study by

Andrews and Lo (2012) showed that vocabulary size scores pronounced a compelling effect of lexical inhibition. That is, the results from the study implied that vocabulary size should correlated with the lexical difficulty and reduced frequency effects. According to Diependaele et al. (2013), LexTALE provides more accurate proficiency measures with higher validity than previously used questionnaire measures.

Over time, spelling has been favored to assess oral proficiency. Even if acknowledging that oral representations are vital for spelling skills, it is easier to measure spelling in silent reading by simply asking the reader to say if a word is correctly spelled or not. If the reader has required the skills necessary for spelling in second language then, a simple task just as the one mentioned above, should be more than satisfying to assess his/her proficiency in L2. However, there are in educational research many other tests that can equally assess spelling but to a deeper degree. An article by Kohnen et al. (2009) reviewed several spelling tests, from orthographic to oral assessment. It was concluded that some tests are better than other and that for L2 learners, oral formatted tests may be better to make sure they know the oral representation of written words. Nonetheless, the scholars recommend spelling tests that contain regular and irregular words to measure orthographic knowledge.

Overall, the presented findings acknowledge that vocabulary and spelling can assess L2 proficiency to a certain extend. However, the complexity of reading processes constitutes some limitations to individually assess how linguistic skills can help predict the level of fluency in second language.

2.4. Eye tracking research in reading

Research on rate and effort associate with reading, such as the measurement of looking time per word, can be collected through eye tracking.

Eye tracking is a research tool with the ability of recording eye movements in the point of gaze (Holmqvist et al., 2011). In other words, the device can follow the pupil of the participant and report specific data about what and for how long a person is looking at a target. The target can be images, words, full texts, among others depending of the subject of interest of research (Holmqvist et al., 2011).

Eye tracking is considered a growing research tool in several fields. To collect data for eye tracking research is required a high-speed video camera that is connected to a display desktop (Rayner et al., 2016). However, the same eye tracker device is not adequate for all experiments and it is usually selected in accordance with laboratory

settings and research purpose. According to Holmqvist et al. (2011) the laboratory should be a sound and light isolated room that prevents the participants' risk of distraction from the experimental task. Thus, infrared light, such as sunlight, results in most cases in complete data loss and should be avoid (Holmqvist et al., 2011).
Besides, working with eye-tracking devices requires methodological skills, competence and training. It is necessary to learn how to work with the device and deal with complications relate with eye-tracking recording (Holmqvist et al., 2011). Very often these complications result from reflections of light in the human eye. Another step to assure good data quality, besides knowing how to operate the device, is setting up the eye camera for each participant (Holmqvist et al., 2011). Such process includes checking the participants eye to identify mascara, contact lenses and difficult glasses (bifocal). Mascara tends to obscure pupil detection while contact lenses and glasses cause multiple reflections in the eye (Holmqvist et al., 2011).

2.4.1. Eye Movements in Reading

The eye takes in light through the pupil that is projected to the retina, which is very sensitive to light (Holmqvist et al., 2011). In the center of the retina, there is a small area called the fovea. According to Schotter and Rayner (2012) the fovea is the region of highest acuity in the eye and only comprises 2 degrees of visual angle. Outside the fovea there is the parafovea, where the acuity decreases significantly. That is why during reading, in order to see a word in a text, we have to move our eyes so that the light from the word gets in the fovea (Holmqvist et al., 2011). The parafoveal processing is the reason why eye fixations during reading are so short (Schotter & Rayner, 2012). That is to say, words can be slightly available to the reader before it is fixated because it is present in the parafovea when the eye is preceding words (Schotter & Rayner, 2012).
Another important part of the human eye is the cornea. The cornea is the front part of the eye that reflects light and protects the pupil (Holmqvist et al., 2011). According to Holmqvist et al. (2011), when recording eye movements, it is ideal to have only one type of reflection, the cornel reflection which is normally the brightest. Furthermore, the corneal reflection offers and additional reference point to pupil-tracking which is helpful to compensate for minor head movements (Holmqvist et al., 2011).
Eye movements, also called saccades, are extremely fast (between 25 to 35 milliseconds) according to the length of the movement (Schotter & Rayner, 2012).

Thus, saccades allow the reader to move the fovea from word to word with higher efficiency (Rayner et.al, 2016).

If the eyes are considerably stable, that is called fixation. It is during fixations that information is received from the words the individual is reading. That is because the vision gets suppressed when the eye is moving from word to word (saccades), resembling "temporary blindness". Fixations are the most reported event in eye-tracking data, even if does not actually relate to a movement since the eye is in fact "still" (Holmqvist et al., 2011, p.21).

While reading the eye has the capacity of picking up information from more than just the currently fixated word (Heister et al., 2012). Parafoveal visual properties make it possible to cover area from 4 characters to the left side of the eye, to a maximum of 15 characters to the right of the current fixation location which can also influence fixation duration. Therefore, during reading of sentences, the fixation duration depends of variables such as predictability and plausibility of words from prior text (Heister et al., 2012).

According to Rayner at al. (2016), about 30 % of the times, readers skip a word to go to the following one. However, just because a work is skipped it does not necessarily mean the word was not processed. Word skipping theories have shown that it can happen due to predictability of words and usually the word can be partially recognized according to context (Rayner et al., 2016).

It can happen that the eyes do not move along to the next word. Some words may require more than one fixation, that is called refixation. Refixations typically occur for long words (in English about 7 or more letters). However, if a reader takes a "step back" in previous words s/he has done a regression (Rayner et al., 2016). Regressions occur from right-to-left in alphabetic writing systems, that is in the opposite direction of reading. Regressions are important in reading studies with eye-tracking, because they can represent failure in comprehension (Rayner et al., 2016).

2.4.2. Eye tracking measures

There are five different types of eye tracking measures that generate data for quantitative research (Holmqvist et al., 2011). These are movement, position, numerosity, latency and distance measures. Movement measures in eye tracking relate to properties of movements that occur during a specific time. Position measures refer to where and for how long a participant has or has not been looking at something.

Frequently, position measures consider duration of fixation ad a certain position (Holmqvist et al., 2011). Numerosity measures are different from the other measures because they account more than one single event (Holmqvist et al., 2011). As the name describes, numerosity measures considers counting the eye movement events. Lastly, latency (measure of time delay) and distance (from one point to another) measures consider spatial distance a temporal latency values of single movements in eye tracking (Holmqvist et al., 2011).

2.4.3. The Areas of Interest (AOI)

The area of interest (AOI) corresponds to the regions around the stimulus. For instance, is often used to investigate if the participant looked where it was expected, if words were skipped, and what properties of their eye movements in the area were looked at. Figure 1 illustrates how the AOI of words is defined.

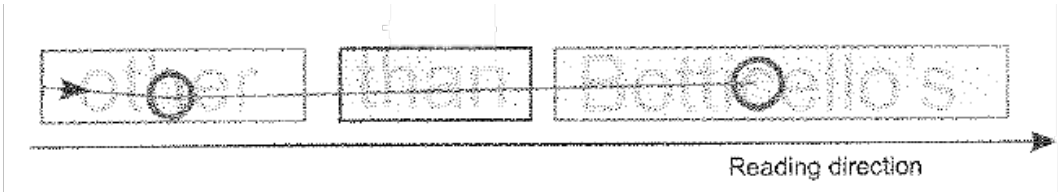

Figure 1. Eye tracking data example of an AOI (word) in a passage: the circles represent the point of gaze and the line represents the path of the eye (from Holmqvist et al., 2011, p.191).

As the example shows (Figure 1), the AOI in words takes a rectangular shape surrounding the target/stimulus. According to previous studies, the amount of time spend on AOI is often related to fixation time (Holmqvist et al., 2011). AOI is not considered to measure average looking time because if the reader looks at a point above the rectangular area the time spend in the fixation is not considered. Because the intent is to study the total effort associated in each passage, the duration of looking points outside the AOI were considered. In other words, looking time in reading also considers the time spent looking at other points outside the AOI of a word.

2.4.4. Eye tracking properties on fixation

The measurements related to how long readers fixate a word are usually fixation duration but also first fixation and single fixation. First fixation duration is the duration of the first fixation on a word. Single fixation duration (also known as SFD) corresponds to the duration of fixation on a word in cases where the word was only fixated once (Schotter & Rayner, 2012). SFD do not including regressive fixations.

Fixation duration is the sum-up of all fixations through each AOI (Holmqvist, et al.2011). Fixation duration can also be named dwell-time, gaze duration, fixation time or total fixation duration and is considered a simple position measurement and the most used in eye-tracking research. Research has found that the fixation duration per word in a text is expected to be more or less 200 milliseconds per word (ms/word) when the eye enters all the AOI in a text. On the other hand, lower values are associated with more superficial reading (Holmqvist et al., 2011). However, raw data on fixation duration always consider the times an individual “loses” eye contact with the screening device. An example of such non-fixation is blinking. Considerable amounts of non-fixations relate to non-processing time (Holmqvist et al., 2011). In other words, longer measures of fixation duration suggest poor situation awareness, uncertainty and difficulty in extracting and work through information. Holmqvist et al. (2011) mentioned that lexical activation and recognition should be measured with first fixation duration since slower cognitive processes affect total fixation duration.

Words that are refixated considers words that are looked at more than once before the reader moves on to the next word. If words are skipped is means that the reader did not fixate the AOI of the word. Some skipped words (e.g. *the, of, in, and)* give good reasons to believe that even when skipped, the eye still processes the information (Schotter & Rayner, 2012).

As previously mentioned, the measures used for this thesis were mean total viewing time per word in both native and second language. The measure was chosen because it is presumably a better index of the total effort associated with each word. Looking time per word is a "late" measure that includes all viewings. More specifically, includes the amount time spent per word, including the time after any regressions and refixations. Even if this measure is similar to total fixation duration it is not to be confused with it, as looking time does not include the AOI of each word.

3. RESEARCH METHODOLOGY

This section will describe in detail the design for the study at hand. After presenting the protocol used for the study, each relevant variable is described individually.
The data collected for this thesis was conducted under MECO: The Multi-lingual Eye-tracking Corpus. MECO is an international cross-linguistic project with a goal of creating a publicly available research database with eye-tracking, comprehension and individual differences data at the trial level for L1 and L2. Information from the project can be found in the following link: https://osf.io/z2nm3/.

3.1. Research Design

The purpose of this study is to investigate the relationship between English viewing time per word with Norwegian viewing time per word and English proficiency scores in Norwegians with age between 19 and 30. This thesis has a quantitative approach with a non-experimental design since independent variables were not manipulated (Tabachnick & Fidell, 2013). That is, there was no control over the outcome results since the participants were not exposed to any kind of intervention or treatment.
Furthermore, this study is descriptive since it measured the effect of two variables on another variable (Tabachnick & Fidell, 2013).
The data collected for this project was collected by three master's students under the on-going project "Eye movements when reading in native and second language". The project consists of a one-time experimental session, with two different parts. The first one involves reading texts in one's native language (L1), while the second part engages reading texts in English (L2). This study considers L1 as Norwegian (Bokmål).
Individual differences were collected in both the L1 and the L2 parts of the study. In order to check for proficiency in one's second language we tested each participant for vocabulary and spelling knowledge. Additional measures were collected during the testing sessions, which are not relevant for the present thesis and will not be reported here.
The statistical analysis of the study was carried out using Jamovi (Version 1.1.3.0.).

3.2. Sample

The three master students (including myself) participating in the project were responsible to recruit the sample. All participants had to have Norwegian as their mother-tongue. Participants were recruited through the student's social circle. The same individuals were asked to complete both parts of the study.

Fifty-seven participants were recruited. Of those, seven participants had to be excluded, giving us a final number of fifty participants for the project. One participant was excluded because results of one variable were lost, five were excluded due to technical problems with the eye-tracking device and the last one had lived in an English-speaking country for more than six months.

According to the protocol from the cross-linguistic project, the target age range was between 17 to 30 years old. The L2 (English-based) study is designed for relatively proficient non-native speakers of English. Therefore, participants with higher levels of English proficiency were preferentially selected. Thus, information regarding education background was taken into consideration and participants that were enrolled in the university of Oslo were preferably chosen.

Fully bilingual participants were excluded, as well as individuals who have lived more than 6 months in an English-speaking country. Other exclusion criteria included contact lenses and bi-focal glasses (problematic for eye-tracking; Holmqvist et al., 2011) or a self-reported diagnosis of a reading, learning or auditory disability. However, of the fifty participants for the study, four opted to use glasses in both experiments.

3.3. Equipment

The data collection of the project recorded movements from an eye-tracking device. The device was located at the Faculty of Educational Sciences, University of Oslo. The device used for the study was SR Research EyeLink 1000Plus. The equipment includes a Host PC, that controls the camera, and a Display PC that presents the experimental stimuli/task.

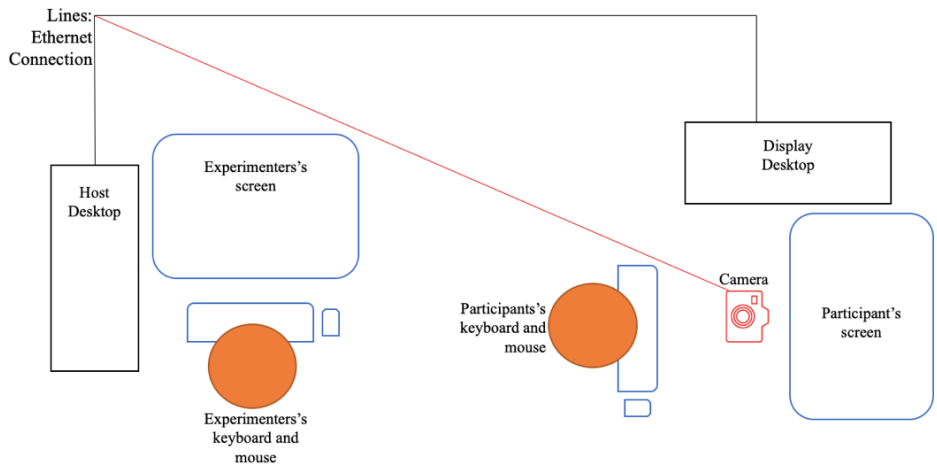

Figure 2. Display of the laboratorial settings

3.4. Procedure

The duration of the combined L1 and L2 study was about 2 hours per participant, with approximately 50 minutes for each part and a 5-10-minute break in between. Participants had to read twenty-four passages in total, twelve texts in their L1 (Norwegian) plus twelve texts in L2 (English). Readers were instructed to read each text silently for comprehension and answer multiple choice questions to each text. Eye-movements were recorded during text reading.

The experiment started by handing out an agreement form to the participant. They were asked to read through it and state whether they agree with the anonymous collection and publishing of the results. That was followed by a questionnaire adapted from the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007). After that, each participant was asked to sit in front of the device and place their chin in the head support, directly placed in front of the eye-tracking camera. The headrest was placed in a way that the eyes were at 75% height of the display screen.

Before starting to read the L1 texts, in order to make sure the eye was properly tracked, it was necessary to find the eye with the camera, zoom and focus on the eye using the lens. Then, we proceeded to increase pupil threshold and decrease the threshold for the corneal reflection of the eye following the standard instructions for operating the equipment. Thereafter, a nine-point calibration and validation were carried out. During calibration, the eye tracker measured how the eye reflects light while validation provided a gaze deviation for each point (Holmqvist et. al., 2011). Participants were

informed if they needed to move their head from the headrest, both calibration and validation had to be re-done. Participants proceeded to read the passages in L1. After that, participants were asked to move to another table without the equipment and we tested for individual measurements in their L1 skills. A 5-minute break was offered to each participant before proceeding to L2 reading. Once the participants sat down, calibration and validation of the eye was required. Again, the participants were informed about redoing both procedures if they needed to move their head. The participants proceeded to the L2 reading and once that was completed, individual skills in L2 were assessed.

3.5. Passages

This section refers to the English and Norwegian passages used in this study. The passages are listed in Appendix 1 and 2. The length of each passage is considered in Table 1 and Table 2.

3.5.1. Norwegian passages

Three master students worked on the translation and creation of the Norwegian passages. Twelve texts in English were received that had to be translated or adapted into Norwegian (L1). These texts were attached to the protocol for the cross-linguistic study and were taken from Wikipedia entries.

According to the protocol, five of the twelve English texts were translated into Norwegian. The translation had to preserve the number of sentences and the general content of each sentence. The remaining seven texts were not translations from English; therefore, new ones were created to have a similar length and general content as the respective English texts. They had to be between ten to fifteen lines, six to twelve sentences long (Table 1). However, passage number 9 had to be edited to sixteen sentences in order to make sense in the native language. Furthermore, they had to match the English texts on genre, topic and level of complexity. Each text should be followed by four yes/no comprehension questions, to ensure that subjects actually read the texts. The questions were formulated based on the new texts content and edited to fit our non-translated texts. The questions from the translated texts were also translated. It was considered that each text should be easy to be read, in a natural way for the native language rather than the English prototype. In order to do that, several native speakers

of Norwegian were asked to read the translated texts and matching questions multiple times. These people were acquaintances of the master students involved in the project.

Table 1. Topic, number of words and number of sentences for each L1 passage

	Topic	Words	Sentences
Passage 1	Greek mythology	177	10
Passage 2	Heraldry weapons	175	8
Passage 3	Doping in sports	175	9
Passage 4	Tasmanian Tiger	186	10
Passage 5	World Environment Day	155	8
Passage 6	Monocle	146	8
Passage 7	Wine tasting	188	9
Passage 8	Orange juice production	174	11
Passage 9	Honey production	240	16
Passage 10	National flag	177	11
Passage 11	International Association for the Conservation of Nature	170	9
Passage 12	Registration/license plate	147	8

3.5.2. English passages

The English texts were chosen by the administrators of the cross-linguistic study to represent a range of reading complexity (from eighth grade of an English-speaking school to the college level). Each text was supposed to occupy one screen exactly followed by two multiple-choice questions. The target texts and accompanying questions were selected from sample materials of the Reading Comprehension segment of the ACCUPLACER test for colleges (The College Board, 2018, https://accuplacer.collegeboard.org/). The multiple-choice questions were designed to measure text comprehension. Table 2 considers the length of the English passages by number of words and number of sentences.

Table 2. Topic, number of words and number of sentences for each L2 passage

	Topic	Words	Sentences
Passage 1	Samuel Morse's biography	160	6
Passage 2	Leonardo Da Vinci's biography	98	4
Passage 3	Amazon Rainforest	107	7
Passage 4	Howard Gardner's intelligence theory	142	11
Passage 5	Internet	185	10
Passage 6	Benefits of sleep	147	9
Passage 7	Preening	173	11
Passage 8	Hibernation	133	8
Passage 9	Leeches	115	6
Passage 10	File-sharing programs	123	8
Passage 11	Pencil's history	116	6
Passage 12	Technology	146	5

3.6. Variables

This study considered one dependent variable and two independent variables. The dependent variable was mean viewing times per word in L2. The independent variables were mean viewing times per word in L1 and English proficiency scores. The independent variable English proficiency scores considers three different measurements (vocabulary, spelling and lexical decision). All measurements are presented individually.

The mean viewing times per word were a result of summing up all the viewings durations in a passage and dividing them by the number or words of that same passage. The reason to use such variable is because it considers the time that each participant took when looking at the passages even when areas of interest (AOI) of the word were skipped. The variables were rated as milliseconds per word. To assess proficiency in English, online tests of vocabulary, spelling and lexical decision were considered.

A control variable is not part of an experiment, but it is important because it can have an effect on the results. The control variable in this study was years of education.

3.6.1. English average looking time (rate L2)

The raw data from the eye-tracking was processed in popEye (Schroeder, 2019). The mean of viewing times per word was retrieved from popEye, developed by Sascha Schroeder. This program parses the raw data into saccades and fixations, assign fixations to lines and letters, and computes reading measures for different levels of analysis (Schroeder, 2019). The preprocessed data file was uploaded to the OSF (Open Science Framework) archive for MECO and retrieved from the following link: https://osf.io/z2nm3/.

3.6.2. Norwegian average looking time (rate L1)

The average looking time for L1 was also retrieved from the online link for MECO (https://osf.io/z2nm3/). Similarly to the previous variable, it refers to the sum of all viewings durations in the Norwegian passages divided by the number of words in a passage.

3.6.3. Vocabulary

To test receptive vocabulary in English, the vocabulary size test adopted from Nation and Beglar (2007) was used. The test was administered online, and it was part of the protocol for MECO. The original version of the test contains 140 multiple-choice items with ten items from each 1000 words family level (of a total of 14,000 words). Each level corresponds to a degree of frequency of the words, which is how common some words are when compared to others. There are three levels of frequency: low, mid or high. High frequency words (1000-2000) are basic words that usually all English speakers come across in everyday life. Mid-frequency words (3000-9000) are more frequent in deliberate learning. Low frequency words start from the 10,000-word family and correspond to words related to specialized studies or areas that usually only readers with large vocabulary knowledge can identify.

Each question of the test contains a target word followed by a sentence in which this word is used, sometimes used in a different morphological form.

The test used for the cross-linguistic study was adapted to facilitate rapid assessment. The first ten items (that represent 1000 most common words of English) were skipped and all participants started from the 2000-word family level. Additionally, a “stop” rule was established so that the participants would not reply to ten items in the same word family level. They would only move to the next word family level if they committed

less than five errors in those first ten items. The target word was followed by four definitions, as shown in the example in Figure 2.

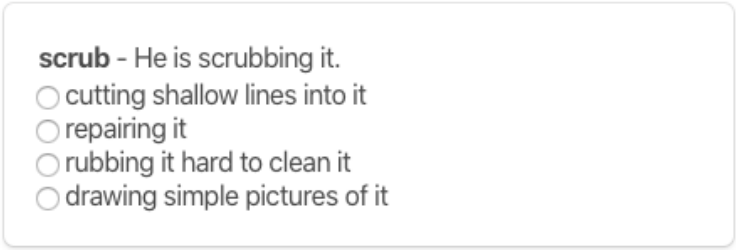

Figure 3. Example of an item from the vocabulary size test

The participants were instructed to choose the option with the closest meaning to the target word in the question. The test was not timed, and it stopped automatically when the participant had more than five errors in a given 1000-word family level or completed the test. The score is the number of correct responses.

3.6.4. Spelling

The spelling test was received from MECO and was ready for online administration. The spelling recognition test used for the study was adopted from Andrews and Hersch (2010). The original test contains 88 English words, including 44 correctly spelled words and the other 44 incorrectly spelled words. Considering the amount of individual differences collected for MECO, the test was adapted for half the original size by the administrators of the cross-linguistic study. Therefore, we assessed spelling for 44 words, of which 22 were misspelled and the other 22 were spelled correctly. Participants had to choose if the word was correctly on incorrectly spelled. The score is the number of correct responses.

3.6.5. LexTALE

Word identification and recognition skills are central to comprehension and reading understanding (Perfetti, Landi & Oakhill, 2005). LexTALE (Lexical Test for Advanced Learners of English, Lemhöfer & Broersma, 2012) is an online lexical decision test design to measure vocabulary knowledge ¨. The test considers British English Spelling. The test consists of 60 words, of which 40 are words and 20 are nonwords. Participants were asked to choose whether the words presented on the screen were real words or not. To reply they had to click on the buttons “yes” or “no”. Each item is between 4-12

letters long. Results were sent to an e-mail address inputted in the beginning of the task. The score is the percentage of correct answers, corrected for the unequal proportion of words and nonwords in the test by averaging the percentages correct for these two item types.

3.6.6. Years of Education

In Norwegian schools, students have English classes for thirteen years, from the age of 6 until the end of high school. After that, if students continue their studies in higher education, their exposure to English tends to increase. That is because major research is usually available in English. Higher exposure to another language is expected to increase the level of proficiency in that language. Therefore, years of education should be considered to influence how well someone would be able to speak and read their L2. The variable years of education was collected through an online version of the LEAP-Q.

3.7. Data Analysis

The data results of this thesis were analyzed in Jamovi Version 1.1.3.0. Jamovi is a very easy to use statistical spreadsheet for basic, intuitive and well-defined statistical analysis (jamovi, 2020).

The statistical analysis started with a descriptive examination of each individual variable in order to analyze the distribution of the values and check for abnormalities. Thereafter, bivariate correlation analysis was carried out to determine the size and direction of the relationship among the continuous variables, including both the dependent and the independent variables. The last step of the analysis was hierarchical multiple regression (Tabachnick & Fidell, 2013) to look at the relationship between the independent variables and the dependent variable.

3.8. Validity and Reliability

Research practices must ensure the reliability and validity of their measurements. The reliability of the measurements used in this study will follow the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments, Mokkink, L. B., et al. 2018; Prinsen et. al., 2018; Terwee et al., 2018) guidelines (Appendix 3). COSMIN's goal is to develop the selection of health measurement instruments in research and clinical practices. COSMIN is a standardized tool meant to assess the methodological quality of studies describing psychometric

properties of a measure (Prinsen et. al., 2018). The guidelines rate the quality of study design and the strength of statistical analyses. Thus, is rated on a four-point scale (inadequate, doubtful, adequate, very good). The instruments used is this study have similar properties and requirements even though they concern different kinds of constructs (i.e., not "health").

The validity of this study follows Cook and Campbell's (1979) validity system for quantitative research. The scholars address four quality requirements, or forms of validity, that assess different aspects of causal relationships (Kleven, 2008). These are construct validity, statistical validity, internal validity and external validity.

3.8.1. Reliability

According to COSMIN (2018), reliability expresses to which degree are the measurements free from measurement error, that is if the results of the study had stable and consistent results. Consider a measurement with different items, such as the spelling test in this study. If in a specific item all participants replied the same, that item did not contribute for the reliability of the test because there is no variance in the results. Reliability is characterized by the proportion of the total variance in the measurements which is due to differences within the sample. Furthermore, it relates to the internal consistency of a measurement. In this study it was considered the internal consistency of each variable in order to observe the relationship among the items. The relationship between items in the measurements was calculated with Cronbach's α (alpha). The table below (Table 3) considers the level of internal consistency according to the result of Cronbach's α.

Table 3. Scores for Cronbach's alpha and level of Internal Consistency (adapted from https://en.wikipedia.org/wiki/Cronbach%27s_alpha)

Cronbach's α	**Internal Consistency**
$0.9 \leq \alpha$	Excellent
$0.8 \leq \alpha < 0.9$	Good
$0.7 \leq \alpha < 0.8$	Acceptable
$0.6 \leq \alpha < 0.7$	Questionable
$0.5 \leq \alpha < 0.6$	Poor
$\alpha < 0.5$	Unacceptable

3.8.2. Validity

According to Kleven (2008) validity is a property of inferences or interpretations. The relevance of various types of validity depend of what kinds of inferences are drawn from the results of a study.

Construct validity refers to whether the measures used in the study really measure the constructs they claim to measure (Gall et al., 2007). Thus, construct validity is about whether the phenomena to be investigated is really measured (Field, 2016). For example, the variable average looking time per word is assumed to be measured through the amount of time spent in a word. The gaze of the participants was observed (to eye tracking resolution), so they were in fact looking. However, it is hard to measure whether they were paying attention to what they were looking at without a comprehension measure. Maybe the participants were thinking of something else or were just bored and scanning the text without deep processing. Nonetheless, the experimental sessions were monitored which should support the eye-tracking properties of average looking time.

Statistical validity is defined by the extent to which valid conclusions can be drawn from the relationship between the dependent variable and the independent variables, so that is statistically significant and strong enough to be of theoretical significance (Lund, 2015). Related to the statistical validity of methods, some thoughts should be introduced. The statistical methods should be adequate to the hypothesis being tested. Hypothesis testing provides a method to reject the null hypothesis (H0) which states there is no effect or relationship among the variables, the alternative hypothesis (H1) says there is an effect between two or more variables. The value of reference to test hypothesis is the p-value (Field, 2016). There are two errors that can occur when checking statistical validity of the study under the null hypothesis. Type I error is when a true null hypothesis is (falsely) rejected and a Type II error is the failure of rejecting a false null hypothesis (Tabachnick & Fidell, 2013).

Internal validity in quantitative methodology relates to the validity of the interpretations that can be taken from observed covariation (Kleven, 2008). In other words, is the extent to which safe conclusions about the relationships between the variables can be drawn.

External validity is defined as the degree to which a causal relationship found in a given study generalizes across diverse settings, measures or persons (Kleven, 2008). Good

external validity should consider a representative sample of the group it is meant to generalize, the selection method, the sample size and the homogeneity of the sample (Field, 2016).

3.9. Research Ethics

The National Committee for Research Ethics in the Social Sciences and the Humanities (NESH) is an impartial advisory body that aims to provide researchers and the research community with recognized norms of research ethics (NESH, 2016).

The project had approval from Norsk Senter for Forskningsdata (NSD), the Norwegian agency responsible for the approval of projects that process personal data. Participants consented to their (unidentifiable, anonymized) data being shared publicly (Appendix 4). That respects the guidelines under paragraph B) Respect for Individuals in NESH's (2016) guidelines. More specifically, covers human dignity, privacy of personal data, duty to inform, consent and obligation to notify, confidentiality of the participants and the re-use of the data collected.

No personal data was collected from the participants. In this sense, names were not associated to the participant and age was only noted as number of years (i.e., no date of birth). Each participant was assigned a computer-generated code to be used across all tasks: enter it in the eye-tracking session, the online session, write it down on the consent form, the language background questionnaire and the offline tasks. The code is a representation of the participants nationality followed by a four-digit number (e.g. NOXXXX). In the participant sheet form it was noted down the age and sex of the participant. However, such information is not relevant for the study at hand.

To avoid spreading computer viruses into the equipment, the storing of data files was done in a new USB memory stick.

4. RESULTS

This section provides the statistical results of the variables of study. Initially, it is presented a descriptive analysis of the variables. Thereafter, each variable is presented individually. Thus, the normality of the variables is assessed by presenting the graphical and statistical results of the individual variables.

The reliability of the variables for the study is considered prior to the descriptive analysis. Thereafter, the correlation analysis of the variables is provided. Lastly, the results for the hierarchical multiple regression analyses are described.

4.1. Descriptive analysis of the individual variables

Table 4 displays the descriptive analysis of the variables. It considers the mean (M), median (Med), standard deviation (SD), skewness, kurtosis and Shapiro-Wilk p-value for the fifty participants (N) from which data was collected.

Table 4. Descriptive statistics of the variables

	N	M	Med	SD	Skewness	Kurtosis	Shapiro-Wilk p
L2 average looking time	50	235	233	62.5	.232	−.080	.996
L1 average looking time	50	218	213	49.6	.247	.585	.738
Vocabulary	50	51.8	48.5	20.7	−.121	−1.040	.005
Spelling	50	32.1	32.0	5.3	−.031	−.480	.760
LexTALE	50	75.6	76.3	11.4	.085	−.660	.725
Years of Education	50	16.1	17.0	1.9	−.785	−.319	.000

Skewness and kurtosis should be considered in order to assess the normality of variables for a smaller sample size (N <100) compared to graphical methods (Tabachnick & Fidell, 2013). Skewness relates to the symmetry of the distribution, while kurtosis defines the peak, that is the tail of the distribution. Preferably, a normally distributed variable should have skewness and kurtosis values of zero. Moreover, positive values for kurtosis indicate that there is an abundance of values in the tails, meaning that the distribution is narrower than the normal distribution. On the other

hand, negative kurtosis values express fewer values in the tails compared to the normal distribution, resulting of thinner tail and a flatter curve than the normal distribution. However, skewness values above 0 indicate that there are too many low values in the distribution so that the curve has a thinner tail on the right side. In such cases we can say the distribution is right skewed. In similar fashion, negative values of skewness indicate that the distribution is left skewed, so there is a higher amount of values to the right (Field, 2016; Tabachnick & Fidell, 2013).

The Shapiro-Wilk test compares the distribution of our data to the normal distribution (Field, 2016). The test is used preferentially for small samples which is the case of this study. Therefore, it is also used to assess the normality of each variable. If the value is less than .05 (the significance value) it means that our data is significantly different from the normal distribution (Field, 2016).

4.1.1. L2 average looking time

The mean for English (L2) looking time was 235 milliseconds per word and standard deviation of 62.5 (Table 3). The variable is approximately normally distrusted since Shapiro-Wilk has a p-value above .05. The variable English average looking time is relatively normal distributed. It shows a skewness value of .232 which indicates a very slight right skew and a kurtosis of −.08.

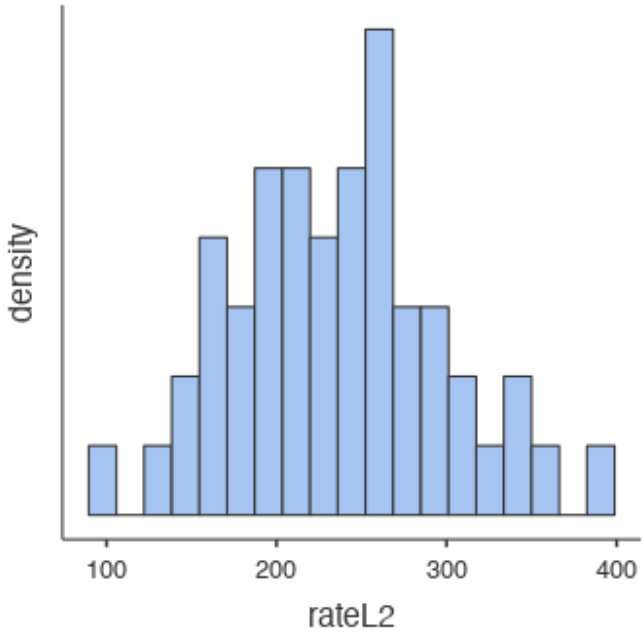

Figure 4. Histogram of mean viewing time per word in English (L2) reading

4.1.2. L1 average looking time

When reading in their native language (Norwegian), the results indicated a mean viewing time of 218 milliseconds per word and a standard deviation of 49.6 (Table 3). The skewness value was .247 and kurtosis was .585. The distribution is a bit irregular but relatively symmetrical considering the difference among the values under the curve. The p-value for Shapiro-Wilk is .738 and therefore we can say the variable is normally distributed.

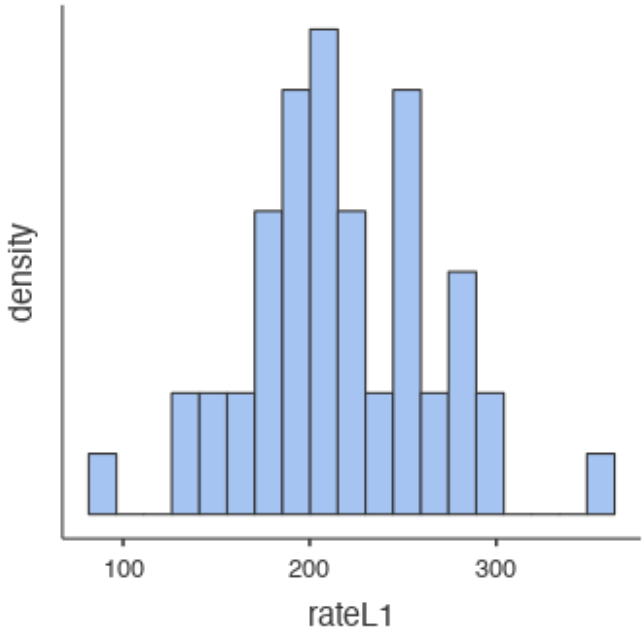

Figure 5. Histogram of mean viewing time per word in Norwegian (L1) reading

4.1.3. Vocabulary

The participants scored a vocabulary assessment in English which varied between 14 and 83, with a mean of 51.8 and a standard deviation of 20.7. The kurtosis value of this variable is −1.040 which corresponds to a somewhat flat curve. Furthermore, it is very slightly left skewed (−.121). Vocabulary assessment in L2 showed some inconsistency in the results and the data is significantly different from the normal distribution (Shapiro-Wilk $p < .05$).

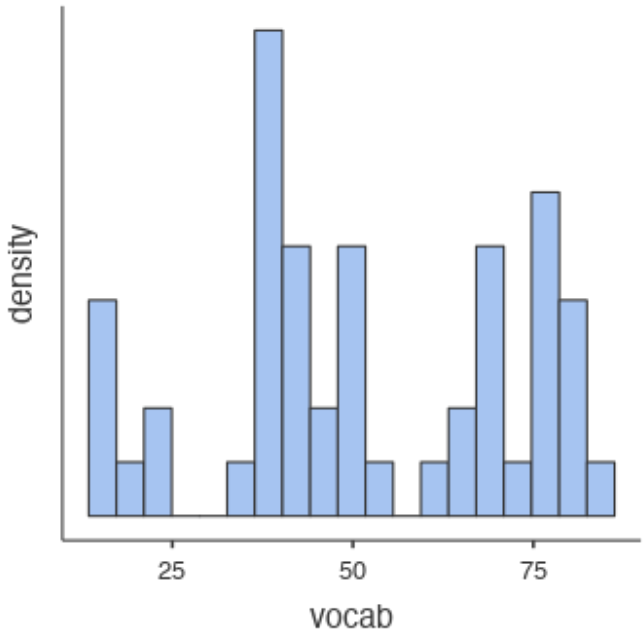

Figure 6. Histogram of English vocabulary scores

4.1.4. Spelling

The spelling test revealed a mean score of 32.1 and standard deviation of 5.28 (Table 3). The values are normally distributed since Shapiro-Wilk $p > .05$. The kurtosis value for the spelling variable was –.480 which means that there exist rather less values in the tails of the distribution. The distribution is essentially symmetrical with a skewness a value of −.031.

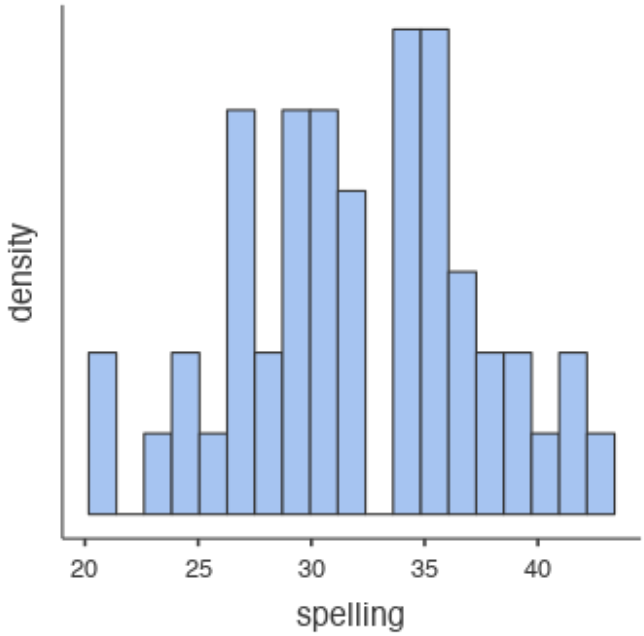

Figure 7. Histogram of English spelling scores

4.1.5. LexTALE

The lexical decision task (LexTALE) scores presented a mean value of 75.6, with standard deviation of 11.4 (Table 3). The p-value for Shapiro-Wilk is above .05 so the values of the variable are normally distributed. In Figure 7, the distribution of LexTALE is fairly symmetrical with a skewness value of .085. However, the curve of the distribution has a somewhat thinner tails resulting of small amount of results in the extremities and a flat curve considering the kurtosis of −.660.

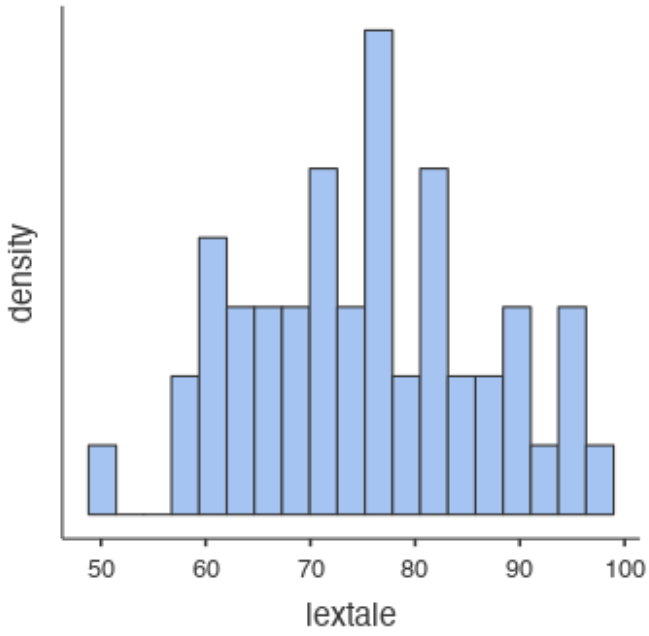

Figure 8. Histogram of scores for lexical decision task (LexTALE) in English

4.1.6. Years of Education

Years of education was the control variable of the study. The histogram show that the range of education is between 11 to 18 years, with mean of 16.1 years and standard deviation of 1.9. The p-value if the Shapiro-Wilk test was less than .05, in other words this control variable is not normally distributed. The distribution is left skewed with a value of –.785. The variable revealed a negative kurtosis of –.319.

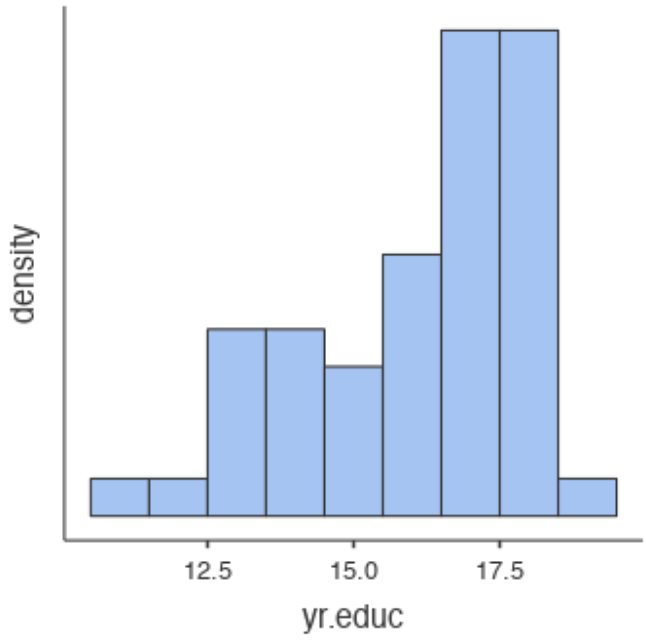

Figure 9. Histogram for years of education

4.2. Reliability

As stated in Chapter 3, the reliability of the individual measurements was done by calculating the Cronbach's α.

The mean viewing time for English and Norwegian was done by considering the viewing time results for the twelve passages in each participant. In both variables (Rate L2 and Rate L1) the measurement was considered to have "excellent" internal consistency (Cronbach's α = .96 in L2 and Cronbach's α = .97 in L1).

Cronbach's α for the spelling test variable was .77, which was the lowest value of the 5 measurements. Even so, it is an "acceptable" consistency between items. Two of the 44 items were excluded to calculate internal consistency because they were too easy to answer, and all participants responded correctly (i.e., there was no variance in these items).

The vocabulary test had 99 items in total, two items showed no variance and did not contribute to the analysis. The reliability of the test was .98 which corresponds to "excellent" instrument to test English vocabulary for this sample.

To measure the internal consistency of LexTALE split-half reliabilities from a previous study were considered. The reason to use results from a previous study is because words and pseudowords in the test behave differently. That is, bias and sensitivity work in opposite directions and therefore Cronbach's α cannot be computed.

Lemhöfer and Broersma (2012) tested in Dutch (72 participants) and Korean (87 participants) speakers of L2 English if LexTALE was a good predictor of vocabulary knowledge. The results for split-half reliabilities was 0.81 for the Dutch group and 0.67 for the Korean participants. The combined score for both nationalities of the study was above 0.87 for both groups. Since Dutch is the language that share more similarities to Norwegian, this study considers the reliability result from the Dutch group (0.81).

4.3. Bivariate Correlation Analysis

As mentioned in Chapter 3, the bivariate correlation analysis aims to calculate the degree of relationship among two or more continuous variables. In statistics, the correlation coefficient can be expressed by Pearson's *r* or Spearman's ρ.
Pearson's *r* is a measure of the strength of a relationship between two continuous variables. Spearman's ρ is a non-parametric measurement based on ranked data (Field, 2016). Both correlations results are reported (Table 4) however only Pearson's *r* will be described.
The correlation coefficient can assume values between –1 and 1. Negative values indicate a negative correlation, in other words higher values in one variable are associated with lower values in the other variable. Hence, a positive correlation, when the correlation coefficient ranges between 0 and +1 means that high values in one variable are correlated with increased values in another variable. If *r* is 0 there is no linear relationship or predictability between the variables. However, to check how much the two variables have in common, r^2 was calculated. The coefficient of determination (r^2 or R^2) is the square of the correlation coefficient and expresses the percentage of variance for two variables (Tabachnick & Fidell, 2013; Field 2016).
Cohen (1992), as cited in Field (2016), suggested that a correlation of .10 is considered a weak correlation, a .30 correlation corresponds to a moderate correlation, while a strong relationship among the variables takes values equal or above .50. Is important to note that in the field of special needs education, research has found that very rarely a strong correlation is found (Gall et al., 2007).
Table 5 displays the correlation among the variables used for this thesis. Table 6 displays the corresponding p-values. The original table of the Bivariate Regression Correlation Matrix from *Jamovi* is attached in Appendix 5 for additional consultation.

Table 5. Results for the Bivariate Correlation Analysis

	L2 Looking time	L1 Looking time	Spelling	Vocabulary	LexTALE	Years of Education
L2 Looking time		.843	.521	.560	.413	−.090
L1 Looking time	.774		.359	.318	.247	−.062
Spelling	.531	.267		.527	.553	−.001
Vocabulary	.568	.248	.508		.617	−.068
LexTALE	.374	.183	.547	.617		.042
Years of Education	−.003	.087	.101	.067	.102	

Pearson's r (listed in black) and Spearman's ρ (listed in blue)

Table 6. p-values for correlation coefficients Pearson's r and Spearman's ρ

	L2 Looking time	L1 Looking time	Spelling	Vocabulary	LexTALE	Years of Education
L2 Looking time		<.001	<.001	<.001	.003	.535
L1 Looking time	<.001		0.010	.024	.084	.671
Spelling	<.001	.061		<.001	<.001	.993
Vocabulary	<.001	.082	<.001		<.001	.638
LexTALE	.007	.205	<.001	<.001		.774
Years of Education	.981	.546	.484	.643	.481	

The significance level for this study is .05. The significance level expresses the probability that we are wrong when rejecting a true null hypothesis (H0). The null hypothesis states that there is no correlation between the variables of study. The p-value can take values between 0 and 1 (Field, 2016).

Considering the results from Table 5, the strongest correlation found was between English and Norwegian mean average viewing time per word. The variables had a positive correlation value of r=0.774 which corresponds to 59.9% of common variance between L1 and L2 mean viewing time per word.

Of the English proficiency scores, vocabulary demonstrated the highest correlation with English looking time per word. Pearson's r value was .568, showing 32.3% of common variance. In the same sense, spelling and L2 viewing time have a common variance of 28.2% and LexTALE has a common variance of 14% to English milliseconds per word. On the contrary, Norwegian reading time did not show significant levels in all proficiency scores in English. Th p-values for the correlations in spelling, vocabulary and LexTALE with L1(Norwegian) looking time are not statistically significant since p

$>$ .05. In other words, the correlation coefficients for English proficiency scores showed a weak relationship with Norwegian looking time per word. Spelling had a common variance of 7.1% (r= .061), vocabulary of 6.2% (r= .082) and LexTALE of 3.3% (r= .205).

Vocabulary and spelling tests had a correlation of .508, which expresses 25.8% of common variance. On a similar note, spelling and LexTALE the correlation coefficient was .547, considering a 29.9% relationship between spelling and lexical knowledge. Lastly, vocabulary and LexTALE had the strongest correlation (r= .617) of the English proficiency scores with 38.1% of variance.

Finally, years of education, the control variable of this study, did not show statistically significant correlation (p > .05) with the other variables.

4.4. Hierarchical Multiple Regression Analysis

As previously stated, the purpose of the study is to investigate whether Norwegian average looking time and English proficiency scores can predict English average looking time.

In Hierarchical Multiple Regression the variables are researcher-selected based on logical and theoretical considerations of the variables. In other words, the researcher selects the order of entry of the variables according to the interest of the study. According to Tabachnick and Fidell (2013), variables that can predict the outcome should be the first to be included in the model. Independent variables can be added separately or in blocks. This type of regression analysis can help predict which combination of variables provide an accurate prediction of a dependent variable. If results from the regression analysis are statistically significant (p < .05 for this study), they are considered to predict scores on the dependent variable based on the independent variable.

In regression analysis, R^2 indicates the proportion of variability in the outcome variable that can be explained by the set of predictor variables. The higher the value, the better the predictor is able to explain the variability in the outcome (Tabachnick & Fidell, 2013).

The mathematical formula for regression states that if X is a predictor for Y and the predicted value (Y') on the dependent variable is calculated by summing the Y intercept (A); the value of Y when all the X values are zero and the slope of the line (B); the change in Y divided by the change in X along with the various independent variables

(*X*) (Tabachnick & Fidell, 2013). The difference between the predicted and the observed values of Y at each value of X represents errors of prediction or residuals. In statistical regression, the goal is to obtain a best-fitting straight line that minimizes the squared errors of prediction which is expressed through the beta coefficient (*B*) mentioned above. B represents the slope of the line and it is the degree of change in the outcome variable for every one-unit of change in the predictor variable. The interpretation is that for every one-unit increase in the predictor value, the outcome will increase (if B > 0) or decrease (if B < 0) by the beta coefficient (Tabachnick & Fidell, 2013).

4.4.1. Results from Hierarchical Multiple Regression Analysis

The coefficient table below (Table 7) considers the contribution of each independent variable to the regression models. The constant expresses the predicted value of the dependent variable when all independent variables are equal to zero for the different models. The first step considers the constant (L2 mean looking time per word) and the control variable, years of education (Model 1). Model 2 expresses the effect of L1 looking time on L2 looking time, after considering years of education. Thus, the next step of the regression model was to check to which extent the English proficiency scores would show an effect on L2 mean viewing time per word. In order to verify this effect, all the proficiency scores (vocabulary + spelling + LexTALE) were placed in another block (Model 3).

Table 7. Coefficients table for the Hierarchical Regression Analysis

	Model	B	SE	t	p	Stand. Estimate
1	(Constant)	237.18	75.43	3.14	.003	
	Years of Education	−.11	4.66	−.39	.980	−.003
2	(Constant)	58.38	52.37	1.12	.271	
	Years of Education	−2.31	2.97	−.78	0.441	−.072
	L1 looking time	.98	.12	8.48	<.001	.781
3	(Constant)	−3.68	47.90	−.08	.939	
	Years of Education	−3.21	2.23	−1.44	.158	−.099
	L1 looking time	.89	.09	9.02	<.001	.649
	Spelling	2.78	1.02	2.73	.009	.235
	Vocabulary	1.02	.27	3.71	<.001	.338
	LexTALE	−.39	.51	−.76	.449	−.071

Model 1: (Constant), years of education
Model 2: (Constant), years of education, L1 looking time
Model 3: (Constant), years of education, L1 looking time, spelling, vocabulary, LexTALE

Considering the results from model 3, Table 7 shows that L1 looking time can significantly predict L2 looking time (B= .89, p < .001). This expresses that when L1 looking time increase by one, the value of L2 looking time will increase by 0.89. Similarly, spelling (B=2.78, p < .009) and vocabulary (B=1.02, p < .001) significantly predicted L2 viewing time. This suggests once vocabulary increases by one, the value of L2 viewing time increases 2.78, and once spelling will increase by one, L2 viewing time per word increases by 1.02. On the contrary, lexTALE (B= −.39, p=.449) and years of education (B=−3.21, p=.158) do not show statistically significance to predict L2 average looking time once all the other predictors are controlled for.
After considering the coefficient table, the equation for regression in this study is: L2 mean viewing time per word = −3.68 + (−3.21*years of education) + (.89*L1 mean viewing time per word) + (2.78*spelling) + (1.02*vocabulary) (−.39*lexTALE).

4.4.2. Assessing models

A well-fitted regression model results in predicted values close to the observed data values. Hence, the squared multiple regression (R^2) is used to measure how well the regression prediction approximate the real data points (Field, 2016). R^2 ranges from 0 to 1 and is the proportion of variance of the dependent variable. When R^2 = 0 it is assumed that the proposed model does not improve prediction over the mean model, while a value of 1 indicates a perfect prediction. Higher values of R^2 the better the predictor is able to explain the variability in the outcome (Tabachnick & Fidell, 2013).

Table 8. Results for the Hierarchical Regression Analysis

Model	R	R²	Adjusted R²	df1	df2	p
1	.003	<.001	−.021	1	48	.981
2	.778	.605	.587	2	47	<.001
3	.891	.793	.769	5	44	<.001

Model 1: (Constant), years of education
Model 2: (Constant), years of education, L1 looking time
Model 3: (Constant), years of education, L1 looking time, spelling, vocabulary, LexTALE

According to the results (Table 8) the proportion of variance of years of education is not statistically significant (p=.981), which reflects on the low value of .3%.
When added L1 looking time per word (Model 2), the native language reading viewing time and years of education have 60.5% proportion of variance in L2 looking time (p < .05).

In Model 3, the variables vocabulary, spelling and lexical decision are included in the previous model. The results showed that the proficiency scores in L2 (English), years of education and L1 viewing time have a variance proportion of 79.3% in L2 looking time ($p < .05$).

Table 9. Model Comparison for the Hierarchical Regression Analysis Models

Model Comparison						
Model	Model	ΔR^2	**F**	**df1**	**df2**	**p**
1	− 2	.604	71.8	1	47	<.001
2	− 3	.189	13.4	3	44	<.001

Table 9 pronounces the comparison between the three regression models. Delta-R-squared (ΔR^2) considers the individual contribution of the individual IVs in each model. By analyzing the results (Table 9), L1 mean viewing time per word contributes with 60.4% for L2 mean viewing time per word prediction. Furthermore, L2 proficiency scores in vocabulary and spelling contribute an additional 18.9% to the prediction model.

4.4.3. Statistical Assumptions of Regression Analysis

The generalization of statistical findings is fundamental to express valid results in research (Field, 2016; Tabachnick & Fidell, 2013). In that sense, five statistical assumptions help generalize a model that can be applied in other samples, beyond the one used in this study. The assumptions are linearity and additivity, homoscedasticity, normality and multicollinearity. The Q-Q plot (Figure 10) and scatterplot (Figure 11) for the residuals are considered to evaluate the assumptions. The residual plots for the regression analysis are listed in Appendix 6 (Q-Q plots) and Appendix 7 (Scatterplots). Linearity refers to the relationship between the independent variables (IVs) and dependent variable (DV) can be characterized by a straight line. Linearity is met when a straight-line is formed in the plot rather than a curved shape (Field, 2016; Tabachnick & Fidell, 2013). Additivity is related to linearity as it is the assumption that the combined effect of several predictors is best described by adding their effects together (Field, 2016). Additivity can only be met if the regression has a linear model. Figure 10 considers the scatterplot of the relationship between each of the IVS and DV. It is clear that the data points fall very close to the regression line and therefore linearity and additivity is satisfied.

Homoscedasticity is related to residual distribution; the spread of the residuals should be fairly constant at each point of the predictor variables. This assumption is met when standardized residuals and standardized predictor values show no pattern (Field, 2016). To assess homoscedasticity, Figure 11 considers the residuals plot of the regression analysis (Model 3). The scatterplot (Figure 11) shows that most of the points do not show a pattern and even if they are not ideally dispersed, the residuals distribution is satisfactory and meets the assumption of homoscedasticity.

Assumptions of normality considers that each variable and all linear combinations of the variables are normally distributed. The errors of prediction (values of the residuals) should be normally distributed around each predicted score (Tabachnick & Fidell, 2013). To assess normality, Figure 11 considers the scatterplot for the residuals.

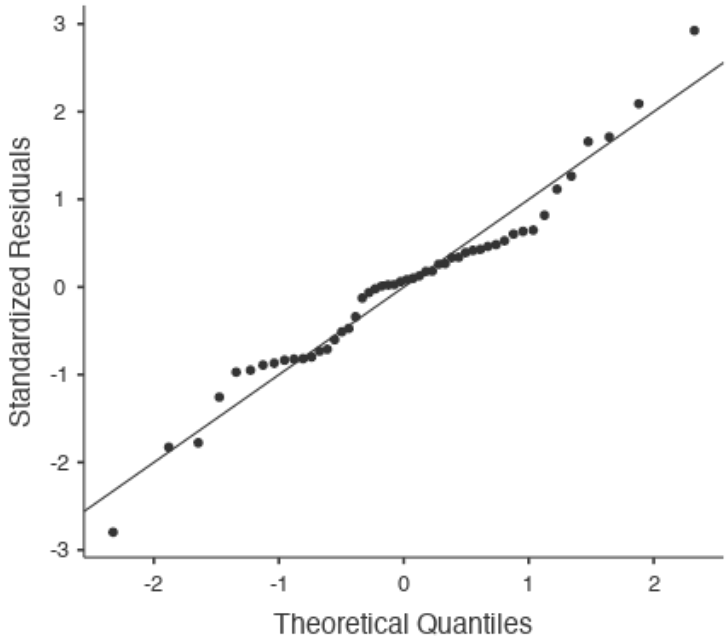

Figure 10. Q-Q plot of Residuals for the Hierarchical Regression Analysis

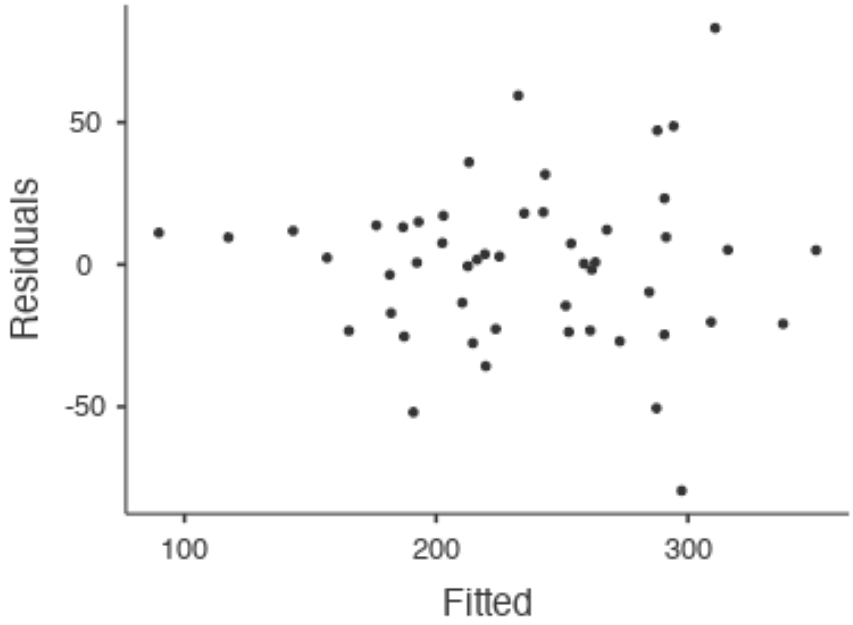

Figure 11. Scatterplot of residuals for the Hierarchical Regression Analysis

The multicollinearity assumption considers the strength of the relationship between two or more predictor variables (Field, 2016). Multicollinearity only exists when there is a strong correlation between two or more prediction variables. High correlations (Pearson's $r > 0.80$) may indicate multicollinearity. Tolerance and variance inflation factors (VIFs) also illustrate if there is multicollinearity between the variables. According to Field (2016) VIFs greater than 10, or tolerances below 0.2, can potentially be problematic. Table 10 presents the results for tolerance and variance inflation factors.

Table 10. Multicollinearity Results

	VIF	**Tolerance**
Years of Education	1.02	.983
L1 looking time	1.10	.908
Spelling	1.58	.635
Vocabulary	1.76	.568
LexTALE	1.84	.544

The results from the bivariate correlation analysis showed that none of the variables reported *r* values above 0.8, which does not give assumptions of multicollinearity. However, considering the collinearity results in Table 10, the VIFs for each variable show a range between 1.02 and 1.84 and tolerance levels are above 0.2.

5. DISCUSSION

This study considered 50 young-adult native readers of Norwegian who volunteered to read twenty-four passages while their eye-movements were recorded. Twelve of these passages were written in the reader's native language (L1), whereas the remaining twelve passages were written in English (L2).

The purpose of this study was to measure to which extent English average viewing time can be predicted. This chapter presents the results of the study after accounting for years of education, reading time in Norwegian and the results of proficiency scores in L2 English in vocabulary and spelling tests.

Thereafter, the reliability and validity of the measurements used for this study are discussed according to COSMIN guidelines (Appendix 3).

5.1. Results of the Study

The results of the study are divided in three sections. The first considers the relationship between average looking time in L1 and L2 and the effect of native language on second language reading time. In the second section, the relationship among the results of the proficiency scores and L2 looking time, and the significance of vocabulary knowledge and spelling on the prediction model are discussed. Lastly, the predictability of L2 looking time is discussed in light of the earlier theoretical background.

5.1.1. The relationship between average looking time in L1 Norwegian and L2 English

The mean for looking time per word in L2 English was 235 milliseconds, whereas L1 Norwegian readers had an average looking time of 218 milliseconds/word. These findings show that L1 reading requires less time to read each word than L2. In other words, Norwegian read faster in their native language than in their second language. The results agree with previous studies on how reading rates in first language are higher when compared to second language (Favreau, et al., 1980; Cop et al., 2015; Brysbaert, 2019). These findings are in line with previous theory on faster reading in L1 (Carver, 1997, as cited in Breznitz, 2006). According to Carver (1997), in native language readers have a faster decoding ability which results on faster reading. Our results support such findings. As expected, smaller values of looking time per word in L1 were reported, and higher values of L2 looking time proved that reading in second language

requires more time (Favreau, et al., 1980; Haynes & Carr,1990; Brysbaert, 2019). However, the length of the passages may have influenced the results. The passages in Norwegian (L1) where considerably larger than the English (L2) passages. The L1 passages had a total of 2110 words, whereas L2 passages had a total of 1645 words. The Norwegian language required more words in order to translate the passages and make meaning of the original version of the texts. Most of the additional words in the Norwegian passages are short which supposedly does not require as much time to read as longer words (Brysbaert, 2019). However, in Norwegian often two words are put together to form a longer word (i.e., *klimaendringer* which translates into *climate change*). This can affect looking time because longer viewings are necessary to read a longer word than shorter one. The difference between looking times in both languages expresses that the reading speed in L1 is faster than in L2, especially because the passages in L2 do not have as much words as L1.

Reading fluency was measured by investigated the effort associated with each word in two languages. That is, it was analyzed how fast Norwegians could read in L1 and L2 in order to measure reading fluency associated with speed. A limitation of this study is that no measures of reading comprehension were taken into account. Such measure would be recommended as it is in line with Carver's (1997) studies on how reading rate require comprehension skills. Overall, it cannot be proved that the readers understand what they are reading, as the participants were selected assuming that they have enough knowledge in both languages to comprehend the passages.

L2 looking time per word and L1 looking time per word had a strong positive correlation (r=.77). L1 looking time had a large effect on L2 looking time as it accounts for 59.9% of the variance. In other words, it means that in this sample, the participants who read faster in Norwegian, were also faster in English.

It cannot be proved that the similarities between Norwegian and English are influencing reading time. However, previous studies that reported strong correlations in L1 and L2 reading were a result of the proximity between writing systems (Haynes & Carr, 1990). The relationship between L1 and L2 in this study is in line with theoretical propositions that native language plays an important role in L2 reading (Olsen, 1990). If so, the results would be in line with previous research which claims the proximity between writing systems in L1 and L2 may facilitate reading (Schepens et al., 2013; Brevik et al., 2016; Haddad, 2019; Verhoeven et al., 2019) and therefore reading rates. Even so,

the disparities in looking time for both languages are not as high when compared to other languages that share the same alphabetic principle (Favreau, et al., 1980).

5.1.2. The relationship between L2 looking time, L2 vocabulary knowledge and spelling

The two tests used to assess vocabulary knowledge were the vocabulary size test and the lexical decision test (LexTALE). The vocabulary size test had 99 questions in total, which also represented the maximum score of the test. For LexTALE, the highest score would be 100, whereas the spelling test had a maximum value score of 44.

The results have shown of this study have shown that the English proficiency scores had a mean of 32.1 for spelling, 51.8 for the vocabulary size test and 75.6 for LexTALE. These results mean that native Norwegians do have a good level of English proficiency, though not very high. These findings are supported by Brevik et al. (2016).

All tests present strong relationship with correlation coefficients (*r*) above .5 among them, which is in accordance with studies that support the use of vocabulary and spelling to assess English proficiency (Kohnen et al.,2009; Lesaux & Siegel, 2003; Lemhöfer & Broersma, 2012).

The correlation between L2 looking time and the spelling test was r=.53 (28% of variance). The vocabulary size had the strongest relation with L2 looking time with r=.57 (32% of variance), while LexTALE r=.37 had a moderate correlation (14% common variance). This means that, of the three proficiency scores, higher vocabulary size values were associated with faster reading in L2. The same effect is noticed with spelling abilities. It is interesting to observe that the decision test did not show a strong correlation with L2 looking time. Since this test also measures vocabulary knowledge the correlation with reading speed would be as strong as the other vocabulary test. However, the results did not support that. The moderate correlation between LexTALE and L2 looking time expresses that higher values of LexTALE are slightly associated with reading speed in English, but not as much as the vocabulary size test. This finding would contradict a previous study that supports the use of lexical decision task to be a better measure of vocabulary knowledge (Diependaele et al., 2013). However, the vocabulary size is a more complex test than the lexical decision test and the correlation coefficient of both tests may be justified by the difficulty level.

The correlations between spelling, vocabulary size and lexical decision with L2 looking time were statistically significant and refer that the level of proficiency in English is

related with reading time in that same language. Overall, L2 proficiency scores accounted for additional 18.9% of the L2 looking time per word, after controlling for L1 looking time, which is in line with the previous research that supports that lexical knowledge is a prime predictor of reading speed in second language (Perfetti, 1985; Breznitz, 2006: Rasinski et al., 2011). Thus, the results have shown that the level of proficiency in L2 does have some influence in L2 reading time.

5.1.3. The prediction of L2 English looking time

The research question of this study considered the predictability of English looking per word, as a second language of native Norwegians. Thus, the study intended to explore the predictability of reading looking time in L2 English when considering years of education, L1 Norwegian reading looking time, vocabulary and spelling proficiency scores.

The sample presented a mean of 16.1 years of education, which means the sample had completed mandatory education of 10 years, upper secondary school, and in most cases three years of higher education. Years of education in Norway did not show statistically significant results to predict looking time per word in English ($p > .05$). The proportion of variance for years of education had the low value of .3% with L2 looking time. Considering that this measure was the control variable of the study, the result was not surprising.

L1 Norwegian looking time contributed with 59.9% to the prediction of L2 English looking time per word. These results have shown that the time spent per word in the readers L1 is very similar to the amount of time the reader spend reading in L2. As earlier mentioned in this chapter, this finding may be due to the close distance between the investigated languages (Brevik et al., 2016). L2 proficiency scores in vocabulary and spelling contributed with additional 18.9% to the prediction model. Lastly, the results of this research have shown that proficiency in English, Norwegian looking time and years of education have a variance proportion of 79.3% in English looking time. This means that to predict L2 English looking time per word, even if the level of knowledge in vocabulary and spelling does add some variance to the prediction model, the time spent per word reading similar passages in Norwegian is more relevant. These results suggest that it seems to be a matter of reading speed in native language, although it is still significant for L2 proficiency.

5.2. Reliability and Validity

As stated in the methodology chapter, reliability and validity are essential to assure quality in scientific research. Research needs to be valid and reliable in order to draw conclusions from the results.

5.2.1. Reliability

Reliability expresses to which degree are the measurements free from measurement error (COSMIN, 2018). Internal consistency of the measurements was assessed on the basis of the reliability coefficient Cronbach's alpha (α). Low levels of reliability are a threat to validity, as it affects the statistical strength of the study (i.e, statistical validity can be weakened by poor reliability). The reliability tests showed that all the variables used for this study had good internal consistency. L2 mean looking time per word had a reliability of α= .96 and L1 mean looking time per word had a reliability of α= .97. Of the proficiency scores, the spelling test had the lowest reliability coefficient value α= .77, the vocabulary size test had a reliability of α= .98 and the lexical decision task (LexTALE) presented a split-half reliability of .81. Overall, the internal consistency of the measurements is not considered a threat for this study.

The risk of bias was assessed with the COSMIN checklist for internal consistency. The checklist is built to limit the possibility of measurement errors. The measurements used for this study met all the relevant points from the checklist. More specifically, all participants had met the sample criteria, the tests were performed on computers to provide similar assessment across participants and in order to reduce possibility of error in administration or scoring The administration of the tests was quite long but it had an appropriate time interval, the administrators made sure the participants were comfortable and were offered to take a break whenever necessary.

Overall, the measurements meet a "very good" level of reliability with a slight disadvantage in the sample size (N=50) as it is ideal to have a sample size superior to 100 (COSMIN, 2018). The items that had to be excluded from the measurements were justified in the results section as they did not add variance to the tests.

5.2.2. Construct validity

Construct validity considers whether the tests used for this study can measure the concept of looking/viewing time and L2 proficiency. According to Kleven (2008), in

educational research, the construct of the measurements is either not measurable or directly visible.

A potential threat to construct validity is the limited time to examine each of the measurements, as they were part of a larger research on eye-tracking movements during reading in L1 and L2. However, the measurements used for this study are considered a strength.

The variables L1 and L2 looking time per word were a result of raw data processed in a program meant to compute eye tracking measures. The readers decided every time when to start and finish each passage by pressing any button in the keyboard. It was not reported that participants had accidentally pressed the keyboard before finishing a passage. While the participants were reading, the administrators had to examine the live recording of eye tracking to make sure the participants were moving along in the text. If some issues happened during recording (i.e, dry eyes or problems with calibration/validation), the administrators where prepared to resolve the matter and improve the recordings when these could be improved. Thus, looking time does describe the average time spent reading each passage.

The proficiency tests in English used in this study were standardized tools that should assure measurement quality. Besides, three measures were used to assess English proficiency, two measures for vocabulary knowledge (receptive vocabulary test and a lexical decision) and one for spelling. These tests were easy to administer as they considered “yes/no” and multiple-choice response format. These types of tests require somewhat low cognitive and attention demands since they do not need require extensive responses. A threat associated with the construct of the proficiency scores in L2 is the "task-impurity" problem. This problem is reported when instruments usually measure more than one variable, since all tasks require multiple skills to be solved (Miyake et al., 2000). This problem is present when vocabulary knowledge was measured. Vocabulary knowledge depends both in the knowledge of the meaning of single words as well as how individual words relate and associate with each other. Besides, different dialects, idioms and expressions reflect vocabulary knowledge. As this study used only individual words to measure L2 vocabulary knowledge, the association and grammar among words was not considered.

Another potential threat of construct validity concerns cross-cultural validity (Kleven, 2008). A possible threat to cross-cultural validity in this study is the translation and adaptation of the Norwegian passages. As stated earlier, L1 looking time is measured

based on the number of words of each passage and the time spent reading them. It was noticed that passages in L1 Norwegian required more words that L2 English to express the same message. Due to the limited time available to collect data, the time spent reading the L2 passages were only tested once by a native Norwegian student participating in MECO. However, the passages were read and edited several times by several native readers of Norwegian, who did not participate in the study, until the texts were considered a good measure to assess fluency. In other words, until the passages were easy to read and had good flow between sentences.

5.2.3. Statistical Validity

As stated in the methodology chapter, statistical validity considers the significance of relationship between the variables and the interpretations of that significance (Lund, 2015).

The results of this study have showed statistically significant correlations between the variables, except for the control variable. If a relationship is found between variables, the possibility to commit a Type I error should be considered. Type I error assumes that there is relationship between the variables, when in fact there is not (Field, 2016). Thus, it is expressed by the interpretation of the significance value. The significance value (.05) used for this study expresses that if our data collection was replicated 100 times, in five occasions it would be expected to find a genuine relationship between the variables, even though there is not (Field, 2016). Overall, it may be concluded that relationships between variables are significant and that there is a small probability of this error to be committed. Type II error is concluding that there is not relationship between variables, even though there is (Field, 2016). In this study the relationship between all variables (apart from the control variable) was statistically significant. Therefore, a Type II error is unlikely to occur.

Similarly, the hierarchical regression model presented significant results among the models. In fact, he strength of the model increased by adding the independent variables. When statistically significant relationship is found between variables, the strength of the relationship should be assessed (Lund, 2015). According to Gall et al. (2007), in educational research the relationship between variables is not expected to have higher correlations than .40. In this study L2 looking time correlated with L1 looking time (r=.77), with the vocabulary size test (r=.57) and with the spelling measure (r=.53). L2 looking time had a strong enough correlation with LexTALE (r=.374). Considering that

a correlation coefficient of .374 is a good enough result in educational research, the correlations between the between the dependent and independent variables is quite strong.

5.2.4. Internal validity

Internal validity is about the extent to which safe conclusions about the relationships between the variables can be drawn. Experimental designs are preferred as threats to internal validity are eliminated automatically. However, in a non-experimental study, such as this one, there is no definitive evidence of causality because the statistical relationship will always have diverse possible causal relationships (Kleven, 2008). In other words, the relationship among variables can be interpreted in different ways and have several explanations. Then, the best way to discuss internal validity is by eliminating possible threats to this study.

As a prediction study, the main goal of this research was to investigate whether the values of one variable can be predicted from scores of other variables (Tabachnick & Fidell, 2013). To do so, the correlation between the independent variables and the dependent variable and the size of common variance were investigated. More specifically, it was investigated whether reading time in native language (L1) and proficiency scores in second language (L2) influence L2 reading time. 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, 2008). For example, proficiency in L2 English can affect L2 reading time and reflect faster reading but it may not be the only explanation for such effect. Besides, it is not possible to draw certain conclusions about the direction of the relationship (Kleven, 2008).

5.2.5. External validity

External validity requires the generalization and transferability of the investigated inferences into wider or other contexts (Kleven, 2008). In order to generalize, the study's sample should meet several requirements. A randomized sampling would be the most appropriate for representativeness. Randomized 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).

The population of this study is people that have Norwegian as their native language (L1) and English as a second language (L2) with various socioeconomic statuses and levels of education. This includes the majority of Norwegians that went to a school and were taught English as a compulsory subject for at least ten years (KD, 2013) and the level of English proficiency among Norwegians has long been high (Brevik et al., 2016). The sample of this study was young-adults (19 to 30 year sold) that had Norwegian as their only native-language and were proficient in English. The sample had different socioeconomic status, different number of years of education and different educational and professional backgrounds, and were not diagnosed with learning, audio or visual disabilities.

According to Gall et al., (2007) a study where correlations between variables are investigated, such as this one, a minimum of 30 participants is desirable to assure the generalizability of the results. A randomized sampling would be the most appropriate for representativeness. Randomized 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). The sample used for this study may not representative of the Norwegian population since it considers only young adults that live or work in Oslo and happen to be acquaintances of three master students of the Department of Special Needs Education, University of Oslo. However, it can be representative for Norwegian young adults in larger cities.

Another threat to external validity is the sample size. The bigger the size of the sample, the more stable the estimates become and the more the uncertainty about the variables decreases (Field, 2016). According to Gall et al. (2007) a minimum of 30 participants is desirable in correlation studies such as this one. The sample of this study had 50 participants which is considered a small sample.

The generalizability of this study can be affected by the texts used to measure looking time in L1 and L2. The texts were considered Norwegian Bokmål and in Norway there are two dialects used across the country. These differences can affect the external validity if the sample is not used to read in Bokmål dialect. Another concern regarding the texts is the difficulty level of the texts. 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. More complex texts may affect the time needed to read

5.3. Summary

The discussion showed that the results of this study support previous evidence for native language and level of proficiency in predicting second language reading rate. Findings from the present study have shown that L2 looking time in reading and L1 looking time are in line with previous findings on how the construct of native language influence second language reading. In this assessment, the level of knowledge in second language should also be considered. In previous studies there were indications that reading comprehension is necessary to assess fluency. However, this study did not considered comprehension as a variable and therefore it was not possible to assess if the effort associated with each word considered successful reading fluency. Nonetheless, the results from the proficiency scores showed that Norwegians have a sufficient knowledge of English to understand the passages. Higher years of education did not reflect faster looking time in L2.

6. CONCLUSION

This study meant to investigate if English looking time per word, as a second language of native speakers of Norwegian, can be predicted by looking time per word in native language and the reader's levels of proficiency in vocabulary and spelling in L2. In the introduction section the following research question was presented: *Can Norwegian looking time per word and English proficiency scores in vocabulary and spelling predict reading looking time in English?*

Results drawn from this study proved that L2 English looking time per word can be predicted according to looking time in L1 Norwegian and proficiency in L2. The results showed that, of the two independent variables, Norwegian reading time was the best predictor of English looking time per word, representing greater variance than the other variable. This significance is possibly a result of the cross-linguistic transfer between both language as similarities between both languages have shown to facilitate the reading process between L1 and L2. Furthermore, it represents that the participants who read fast in native language where also fast readers in English. Vocabulary knowledge and spelling were significant to predict L2 reading whereas years of education was not. Proficiency in L2 does help predict the reading ability in L2 viewing time per word. This conclusion supports previous findings on how higher knowledge in a language affect reading speed.

6.1. Limitations of the study

At this point, several limitations of the present study should be acknowledged. A significant methodological limitation of this study is the sample size. Smaller sample sizes do not give a good view of a phenomenon in the population and therefore a larger number would increase the validity of this study. Similarly, the selection of the participants was a very challenging task that limited the sample size.

Another limitation of this study was not adding a measurement of comprehension to the analysis. Reading comprehension would give this research a better understanding on reading fluency.

Lastly, it is considered a limitation of this research the lack of studies that considered the variable looking time per word. Even if this variable does represent the effort associated with each word, would have been better to use a more common reading rate variable in eye tracking research. However, due to complications in the collection of

data with eye-tracking recording, looking time per word fitted better the purpose of the study.

6.2. Implications in Pedagogical and Educational practices

As reading and writing disabilities seem to have a great impact in Norway, studies in L2 English and L1 Norwegian reading may help understand the linguistic transfer between these two languages and consequently help students with special needs education work on strategies that improve reading in L2. Besides, understanding the nature of cross-language transfer is of great importance for education as such information provides insight into how language learning can be facilitated.

6.3. Potential follow-up and future research

Educational research tends to focus on native language due to the complex process of learning how to read. The results of this study showed that the values of looking time in L1 and L2 were very close, that is native readers of Norwegians who read fast, read also fast in English. The relationship between reading speed in L1 and L2 can be attributed to the similarity between the studied languages. To better understand this process, future research on reading speed across native and second language is recommended.
The explained variance in L2 prediction by L1 looking time also requires further research. In that sense, studies on similarities between Norwegian and English are suggested. Equally important, research on L1 and L2 reading may help understand the processes involved in second language acquisition, which may improve learning practices for pupils with special needs education.

REFERENCES

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General*, *139*(2), 299.

Andrews, S., & Lo, S. (2012). Not all skilled readers have cracked the code: Individual differences in masked form priming. Journal of Experimental Psychology: Learning, Memory, and Cognition, 38, 152–163. doi:10.1037/a0024953

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

Breznitz, Z. (2006). *Fluency in reading: Synchronization of processes*. Mahwah, N.J: L. Erlbaum Associates.

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, *109*, 104047. https://doi.org/10.1016/j.jml.2019.104047

Cop, U., Drieghe. D., Duyck, W. (2015). *Eye Movement Patterns in Natural Reading: A Comparison of Monolingual and Bilingual Reading of a Novel*. PLoS ONE 10(8): e0134008. doi:10.1371/journal.pone.0134008

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally.

Cummins, J. (1979). Linguistic Interdependence and the Educational Development of Bilingual Children. *Review of Educational Research*, *49*(2), 222–251. https://doi.org/10.3102/00346543049002222.

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first and second language word recognition: A lexical entrenchment account. Quarterly Journal of Experimental DOI: 10.1080/17470218.2012.720994

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading research quarterly*, *38*(1), 78-103. https://doi.org/10.1598/RRQ.38.1.4

Enander, H., Weibull, J., Christensen, J., Sandvik, G., & Joys, C. (2020). Encyclopædia Britannica, inc. Retrieved May 15, 2020 from https://www.britannica.com/place/Norway.

Favreau, M., Komoda, M. K., & Segalowitz, N. (1980). Second language reading:

Implications of the word superiority effect in skilled bilinguals. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *34*(4), 370.

Field, A. (2016). *An Adventure in Statistics: The reality enigma* (1st ed.). Los Angeles: SAGE.

Frederickson, N. & Cline, T. (2015). *Special Educational Needs Inclusion and Diversity*. (pp. 322-396). Maidenhead: McGraw-Hill Open University Press.

Fraser, C. A. (2007). Reading rate in L1 Mandarin Chinese and L2 English across five reading tasks. *Modern Language Journal, 91:* 372-394.

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

Green, K., Tønnessen, F.E., Tambs, K., Thoresen, M. & Bjertness, E. (2009). Dyslexia: Group Screening among 15–16-Year-Olds. in *Scandinavian Journal of Educational Research*, 53:3, 217-227, DOI: 10.1080/00313830902917246

Gough, P., & Tunmer, W. (1986). Decoding, reading and reading disability. Remedial and Special Education, 7(1), 6-10. http://dx.doi.org/10.1177/074193258600700104

Guo, Y., & Roehrig, A. D. (2011). Roles of General versus Second Language (L2) Knowledge in L2 Reading Comprehension. *Reading in a foreign language*, *23*(1), 42-64.

Haddad, E. (2019). What is phonological awareness in L2?. *Journal of Neurolinguistics* (50). 17-27.

Haynes, M., & Carr, T. H. (1990). *Writing system background and second language reading: A component skills analysis of English reading by native speaker-readers of Chinese.* In T. H. Carr & B. A. Levy (Eds.), *Reading and its development: Component skills approaches* (p. 375–421). Academic Press.

Harbert, W. (2007). *The Germanic Languages*. Cambridge Language Surveys. Cambridge University Press. DOI: 10.1016/B0-08-044854-2/02189-1 . Retrieved 22 May 2020.

Heister, J., Würzner, K.M., & Kliegl, R. (2012). Analyzing large datasets of eye movements during reading. In J. S. Adelman (Ed.), *Current issues in the psychology of language. Visual word recognition: Meaning and context, individuals and development* (pp. 102–130). Psychology Press.

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

Ibsen, E. B. (2002). *English in Norway*. Retrieved April 23, 2020 from http://www.eva.dk/projekter/2002/ evaluering-af-faget-engelsk-igrundskolen/projektprodukter/assessmentofenglish.pdf

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning,* 64(1), 160-212.

Kleven, T. A. (2008). Validity and validation in qualitative and quantitative research. *Nordic Studies in Education*, *28*(03), 219-233.

Kohnen, S., Nickels, L., & Castles, A. (2009). Assessing spelling skills and strategies: A critique of available resources., Australian Journal of Learning Difficulties, *14:1*, 113-150, DOI: 10.1080/19404150902783450

Kunnskapsdepartementet. (2017, 1st September). *Fag og læreplaner.* https://www.regjeringen.no/no/tema/utdanning/grunnopplaring/artikler/innhold-vurdering-og-struktur/id2356931/

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick valid Lexical Test for Advanced Learners of English. Behavior Research Methods, 44, 325-343. DOI 10.3758/s13428-011-0146-0.

Lesaux, N. K., & Siegel, L. S. (2003). The Development of Reading in Children Who Speak English as a Second Language. *Developmental Psychology, 39*(6), 1005–1019. https://doi.org/10.1037/0012-1649.39.6.1005

Lund, T. (2015). Metodologiske prinsipper og referanserammer. I K. T. A. Lund Thorleif, Kvernbekk Tone, Christophersen Knut-Andreas (Red.), *Innføring i forskningsmetodologi* (2. utg., s. 79-123). Bergen: Fagbokforlaget.

Lyster, S. (2001). Language and Reading – Development and Disorders. In Johnsen, B. H., & Skjørten, M.D. *Education – Special Needs Education, An Introduction.* (pp. 189-202). Oslo: Unipub AS

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*.

McIntosh, C. (2013). *Cambridge advanced learner's dictionary* (Ed.). Cambridge University Press. Retrieved May 15, 2020 from https://dictionary.cambridge.org/

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.

Melby-Lervåg, M., Lervåg, A. (2013). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin*, *,140*(2), 409-433 doi: 10.1037/a0033890.

Moats, L. C. (1983). A comparison of the spelling errors of older dyslexic and second-grade normal children. *Annals of Dyslexia, 33,* 121-140. https://doi.org/10.1007/BF02648000

Mokkink, L. B., De Vet, H. C., Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research*, *27*(5), 1171-1179.

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.

Nergård-Nilssen, T., & Hulme, C. (2014). Developmental dyslexia in adults: Behavioural manifestations and cognitive correlates. *Dyslexia*, *20*(3), 191-207. https://doi.org/10.1002/dys.1477

NESH. (2016). *Guidelines for Research Ethics in the Social Sciences, Humanities, Law and Theology.* Retrieved March 26, 2020, from https://www.etikkom.no/en/ethical-guidelines-for-research/guidelines-for-research-ethics-in-the-social-sciences--humanities-law-and-theology/.

OECD (2013). PISA 2012 Results: Ready to Learn: Students' Engagement, Drive and Self- Beliefs (Volume III). PISA, Paris: OECD Publishing. http://dx.doi.org/10.1787/9789264201170-en

Oller, J. W., & Tullius, J. R. (1973). Reading skills of non-native speakers of English. *International Review of Applied Linguistics*, *11*(1), 69-79.

Olsen, S. (1999). Errors and compensatory strategies: a study of grammar and

vocabulary in texts written by Norwegian learners of English. *System, 27(2),* 191-205. https://doi.org/10.1016/S0346-251X(99)00016-0

Olson, A. M., & Sabers, D. (2008). Standardized tests. *21st century education: A reference handbook*, *1*, 423-430. http://dx.doi.org/10.4135/9781412964012.n46

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific studies of reading*, *11*(4), 357-383. https://doi.org/10.1080/10888430701530730

Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (p. 145–174). Lawrence Erlbaum Associates, Inc. accessed at

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of functional literacy*, *11*, 67-86.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. *The science of reading: A handbook*, 227-247. https://doi.org/10.1080/10888438.2013.827687

Potter, S. & Crystal, D. (2019). *English language.* Encyclopædia Britannica, inc. Retrieved May 15, 2020 from https://www.britannica.com/topic/English-language/Vocabulary

Prinsen, C. A., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C., et al. (2018). COSMIN guideline for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res, accept. 661-683). New York: Taylor & Francis.

Rasinski, T., Samuels, S. J., Hiebert, E., Petscher, Y., & Feller, K. (2011). The relationship between a silent reading fluency instructional protocol on students' reading comprehension and achievement in an urban school setting. *Reading Psychology*, *32*(1), 75-97.

Rayner, K., Schotter, E. R., Masson, M. E., Potter, M. C., & 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.

Santoro, L. E., Coyne, M. D., & Simmons, D. C. (2006). The reading–spelling connection: Developing and evaluating a beginning spelling intervention for children at risk of reading disability. *Learning Disabilities Research & Practice*, *21*(2), 122-133. Doi: 10.1111/j.1540-5826.2006.00212.x

Schmitt, N. & McCarthy, M. (2001). *Vocabulary: Description, acquisition, and pedagogy*. Cambridge: Cambridge University Press.

Schotter, E. R., & Rayner, K. (2012). Eye movements and word recognition during reading. In J. S. Adelman (Ed.), *Current issues in the psychology of language. Visual word recognition: Meaning and context, individuals and development* (p. 73–101). Psychology Press.

Schroeder, S. (2019). *popEye - An R package to analyse eye movement data from reading experiments.* GitHub repository, https://github.com/sascha2schroeder/popEye

Štajner, S., Yaneva, V., Mitkov, R., & Ponzetto, S. P. (2017). Effects of lexical properties on viewing time per word in autistic and neurotypical readers. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 271-281). DOI: 10.18653/v1/W17-5030

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

Terwee, C.B., Prinsen, C.A.C., Chiarotto, A., Westerman, M.J., Patrick, D.L., Alonso, J., Bouter, L.M., de Vet H.C.W., Mokkink LB. (2018). (2018). COSMIN methodology for assessing the content validity of PROMs: User manual.

The jamovi project (2020). *jamovi* (Version 1.2) [Computer Software]. Retrieved from https://www.jamovi.org

Van den Bos, K. P., Zijlstra, B. J., & lutje Spelberg, H. C. (2002). Life-span data on continuous-naming speeds of numbers, letters, colors, and pictured objects, and word-reading speed. *Scientific Studies of Reading*, *6*(1), 25-49.

Verhoeven, L. (2010). Second language reading acquisition. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. Afflerbach (Eds.). *Handbook of Reading Research* (p. 661-683). New York: Taylor & Francis.

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

Vikør, L. (2005). *Fakta om Norsk*. Språkrådet. Retrieved May 29, 2020 https://www.sprakradet.no/Spraka-vare/Norsk/fakta-om-norsk/).

Walter, C. (2004). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Linguistics*, 25, 315–339.

Yamashita, J., & Ichikawa, S. (2010). Examining reading fluency in a foreign language: Effects of text segmentation on L2 readers. *Reading in a Foreign Language*, *22*(2), 263-283.

APPENDIX 1: L1 PASSAGES AND QUESTIONS

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? J
2. Symboliserte Janus nattetid? N
3. Var dørene til Janus tempel åpent i fredstid? N
4. Var Janus en gresk gud? N

2.
Våpen med den likeverdig 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? J
2. Kan skjoldmerke har bare et figurer? N
3. Fiket norsk folk nye våpen etter 1400? N
4. Representere heraldikk våpen et land og nasjonalitet? N

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

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? J
2. Bidro introduksjon av sauer til utryddelsen av pungulven? N
3. Blir pungulven sammenlignet med en tiger på grunn av halen? N
4. Fantes pungulven i Asia? N

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

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? N
2. Ble monokler assosiert med et stereotypisk utseende? J
3. Var stilendring grunnen til at man sluttet å ta i bruk monokler? N
4. Ble monokler brukt av medlemmer av sosial og militær elite? J

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

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

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. N
2. Nå lages kubene for det meste av halm eller tre. N
3. Biene produserer voks selv. J
4. Nå kan man hente ut honning og beholde biene fra år til år. J

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

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. J
2. Organisasjonen er hovedsakelig drevet av frivillige. N
3. Hovedmålet er å fremme naturens bevarelse til allmenheten. N
4. Organisasjonen utga listen med utrydningstruet arter. J

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. J
2. Størrelsen på skiltene var standardisert før andre verdenskrig. N
3. Alle nasjonale kjøretøyregister inkluderer både motorkjøretøy og båter. N
4. Et registreringsnummer for kjøretøy kan være bygd opp av bokstaver, siffer og symboler. N

APPENDIX 2: L2 PASSAGES AND QUESTIONS

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: COSMIN CHECKLISTS FOR RELIABILITY

Retrieved from Mokkink, L. B., et al. 2018; Prinsen et. al., 2018; Terwee et al., 2018

Internal consistency

Like structural validity, internal consistency is only relevant for PROMs based on a reflective model. Furthermore, internal consistency should be assessed for unidimensional (sub)scales. Therefore, unidimensionality or structural validity using e.g. factor analysis should be assessed for each scale or subscale in the study or evidence for structural validity obtained in a previous study in a sample from a comparable target population should be available.

Internal consistency	very good	adequate	doubtful	inadequate	NA	Justification
Design requirements						
1 Check whether a scale or a subscale is unidimensional	Evidence provided that each scale or subscale is unidimensional		Unclear whether each scale or subscale is unidimensional	the scale or subscale is NOT unidimensional		RoB Box 4 (1)
2 Perform the analysis in a sample with an appropriate number of patients (taking into account expected number of missing values)	≥100 patients	50-99 patients	30-49 patients	<30 patients		Sample size
3 Provide a clear description of how missing items will be handled	The way missing items will be handled is clearly described		The way missing items will be handled is not clearly described			Original CC
Statistical methods						
4 For continuous scores: calculate Cronbach's alpha or Omega for each unidimensional scale or subscale	Cronbach's alpha, or Omega will be calculated		Only item-total correlations will be calculated	No Cronbach's alpha and no item-total correlations will be calculated	Not applicable	RoB Box 4 (2)
5 For dichotomous scores: calculate Cronbach's alpha or KR-20 for each unidimensional scale or subscale	Cronbach's alpha or KR-20 will be calculated		Only item-total correlations will be calculated	No Cronbach's alpha or KR-20 and no item-total correlations will be calculated	Not applicable	RoB Box 4 (3)
6 For IRT-based scores: calculate standard error of theta (SE (θ)) or reliability coefficient of estimated latent trait value (index of (subject or item) separation) for each unidimensional scale or subscale	SE(θ) or reliability coefficient will be calculated			SE(θ) or reliability coefficient will NOT be calculated	Not applicable	RoB Box 4 (4)

Box 4. Internal consistency

Does the scale consist of effect indicators, i.e. is it based on a reflective model? [1] yes / no

Design requirements	very good	adequate	doubtful	inadequate	NA
1 Was an internal consistency statistic calculated for each unidimensional scale or subscale separately?	Internal consistency statistic calculated for each unidimensional scale or subscale		Unclear whether scale or sub scale is unidimensional	Internal consistency statistic NOT calculated for each unidimensional scale or sub scale	
Statistical methods					
2 For continuous scores: Was Cronbach's alpha or omega calculated?	Cronbach's alpha, or Omega calculated		Only item-total correlations calculated	No Cronbach's alpha and no item-total correlations calculated	Not applicable
3 For dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	Cronbach's alpha or KR-20 calculated		Only item-total correlations calculated	No Cronbach's alpha or KR-20 and no item-total correlations calculated	Not applicable
4 For IRT-based scores: Was standard error of the theta (SE (θ)) or reliability coefficient of estimated latent trait value (index of (subject or item) separation) calculated?	SE(θ) or reliability coefficient calculated			SE(θ) or reliability coefficient NOT calculated	Not applicable
Other					
5 Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

[1] If the scale is not based on a reflective model, internal consistency is not relevant

Box 6. Reliability

Design requirements	very good	adequate	doubtful	inadequate	NA
1 Were patients stable in the interim period on the construct to be measured?	Evidence provided that patients were stable	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable	
2 Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate or time interval was not stated	Time interval NOT appropriate	
3 Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
Statistical methods					
4 For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated	Not applicable
5 For dichotomous/nominal/ordinal scores: Was kappa calculated?	Kappa calculated			No kappa calculated	Not applicable

Box 7. Measurement error

Design requirements	very good	adequate	doubtful	Inadequate	NA
1 Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable	
2 Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate or time interval was not stated	Time interval NOT appropriate	
3 Were the test conditions similar for the measurements? (e.g. type of administration, environment, instructions)	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
Statistical methods					
4 For continuous scores: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	SEM, SDC, or LoA calculated	Possible to calculate LoA from the data presented		SEM calculated based on Cronbach's alpha, or on SD from another population	Not applicable
5 For dichotomous/nominal/ordinal scores: Was the percentage (positive and negative) agreement calculated?	% positive and negative agreement calculated	% agreement calculated		% agreement not calculated	Not applicable
Other					
6 Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

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 — Areti Kalaitzi, Sara Fonseca, Veronica Tønnesen

Prosjektansvarlig og veileder — *Masterstudenter*

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet «Eye movements in reading Norwegian and English», og har fått anledning til å stille spørsmål. Ved å delta i datainnsamling samtykker jeg til:

- ☐ å delta i lesestudiet hvor øyebevegelsene mine blir registrert
- ☐ å delta i tester og fylle ut spørreskjema knyttet til kognitive, språk- og leseferdigheter.
- ☐ at mine opplysninger behandles anonymt utenfor EU
- ☐ at mine opplysninger lagres anonymt etter prosjektslutt, for videre forskning
- ☐ at mine opplysninger behandles frem til uavgrenset tid

(uten signatur eller registrering av navn)

APPENDIX 5: CORRELATION MATRIX AND P-VALUES FOR PEARSON'S *R* AND SPEARMAN'S *ρ*

Table 11. Bivariate Regression Correlation Matrix

Correlation Matrix

		rateL2	rateL1	spelling	vocab	lextale	yr.educ
rateL2	Pearson's r	—					
	p-value	—					
	Spearman's rho	—					
	p-value	—					
rateL1	Pearson's r	0.774	—				
	p-value	4.26e-11	—				
	Spearman's rho	0.843	—				
	p-value	1.57e-14	—				
spelling	Pearson's r	0.531	0.267	—			
	p-value	7.23e0-5	0.0605	—			
	Spearman's rho	0.521	0.359	—			
	p-value	1.06e0-4	0.0104	—			
vocab	Pearson's r	0.568	0.248	0.508	—		
	p-value	1.69e0-5	0.0822	1.66e-4	—		
	Spearman's rho	0.560	0.318	0.527	—		
	p-value	2.38e0-5	0.0244	8.40e-5	—		
lextale	Pearson's r	0.374	0.183	0.547	0.617	—	
	p-value	0.00741	0.2045	4.01e-5	1.80e-6	—	
	Spearman's rho	0.413	0.247	0.553	0.617	—	
	p-value	0.00286	0.0841	3.17e-5	1.79e-6	—	
yr.educ	Pearson's r	-0.003	0.087	0.101	0.067	0.102	—
	p-value	0.98096	0.5464	0.484	0.643	0.481	—
	Spearman's rho	-0.090	-0.062	-0.001	-0.068	0.042	—
	p-value	0.53516	0.6712	0.993	0.638	0.774	—

APPENDIX 6: Q-Q PLOTS

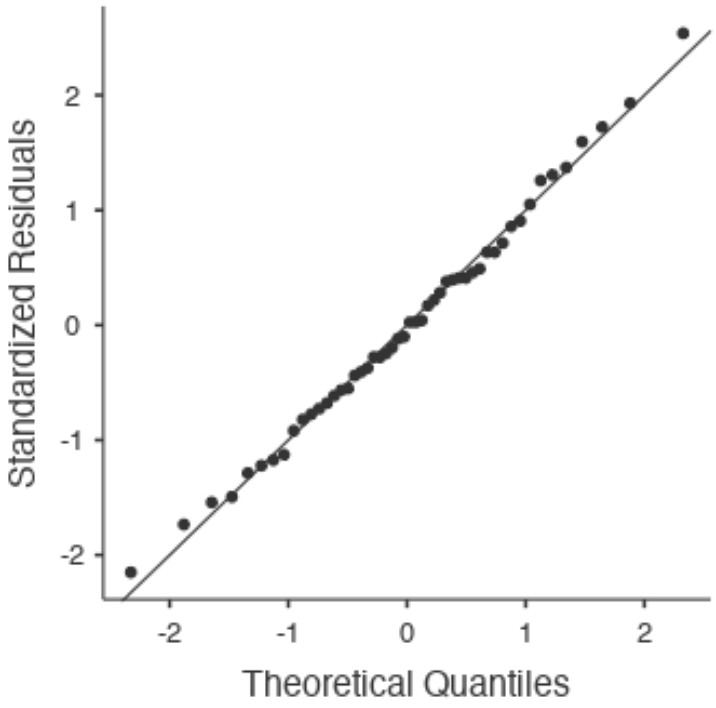

Figure 12. Q-Q plot for L2 looking time per word

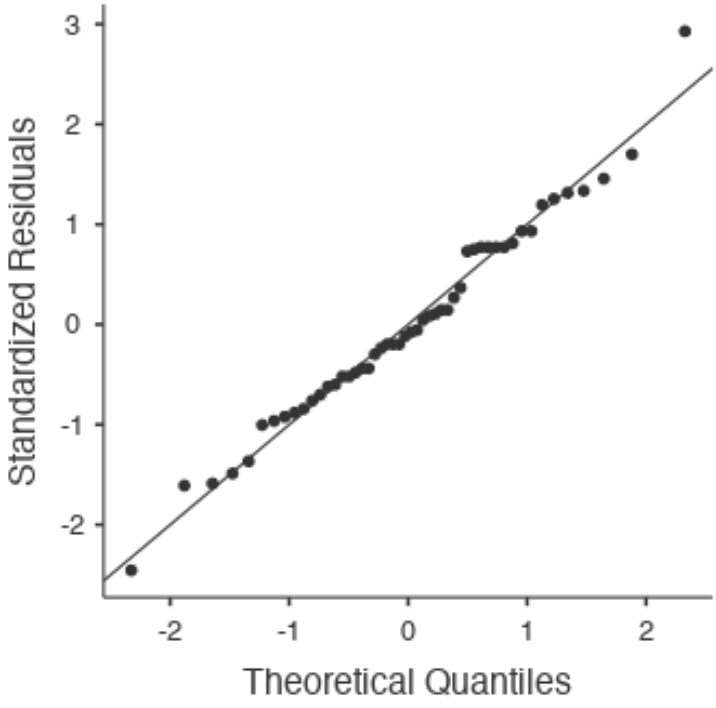

Figure 13. Q-Q plot for L1 looking time per word

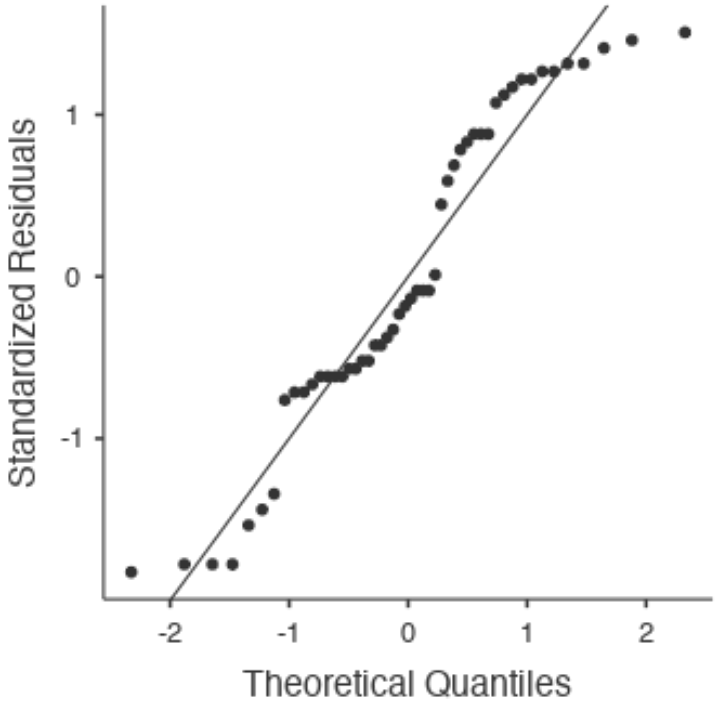

Figure 14. Q-Q plot for the vocabulary size test

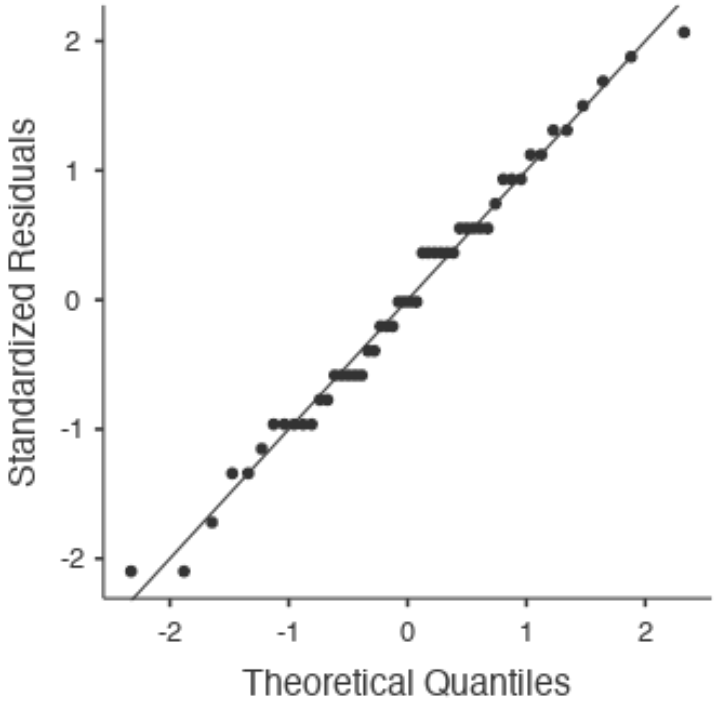

Figure 15. Q-Q plot for the spelling test

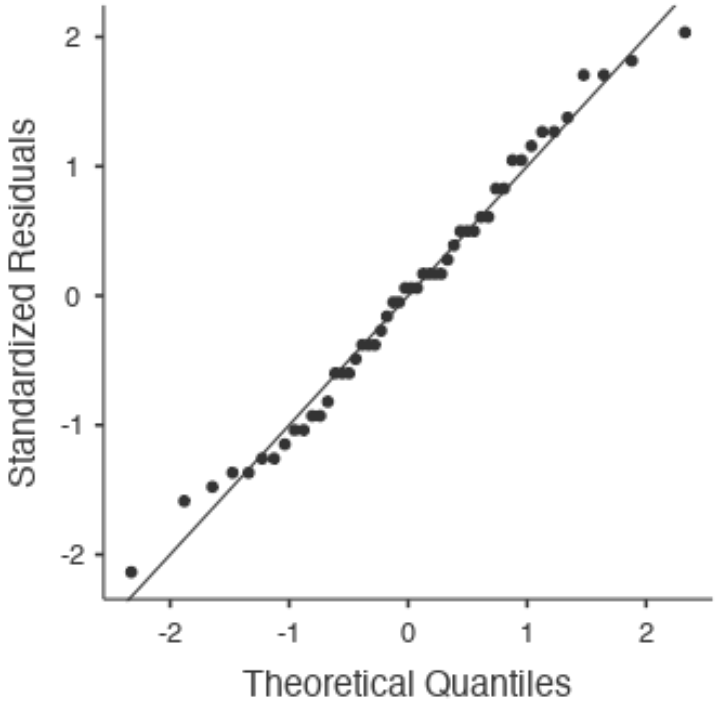

Figure 16. Q-Q plot for the spelling test

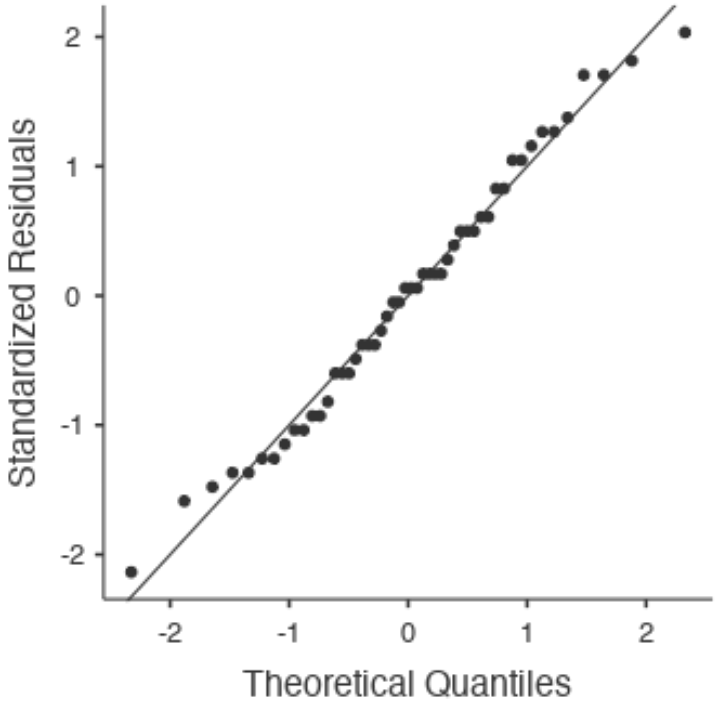

Figure 17. Q-Q plot for the lexical decision test (LexTALE)

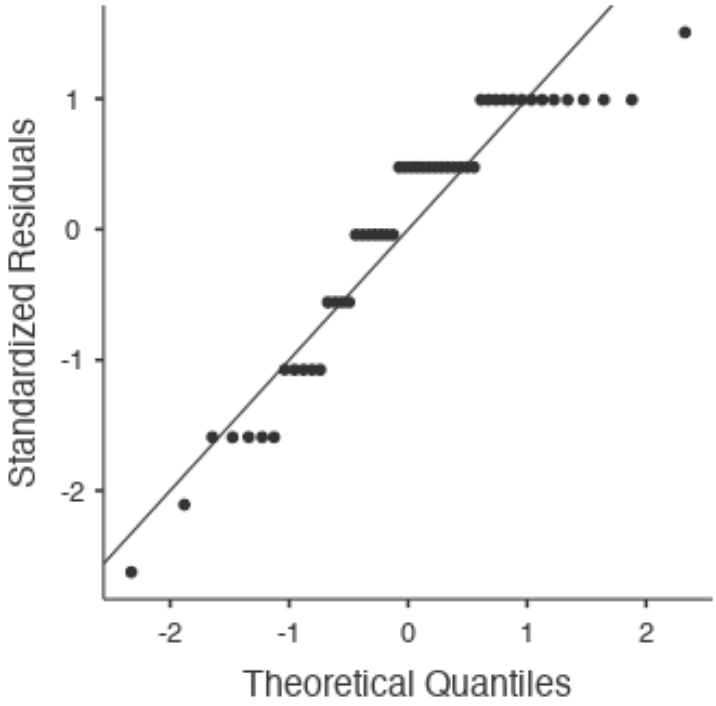

Figure 18. Q-Q plot for years of education

APPENDIX 7: RESIDUAL PLOTS

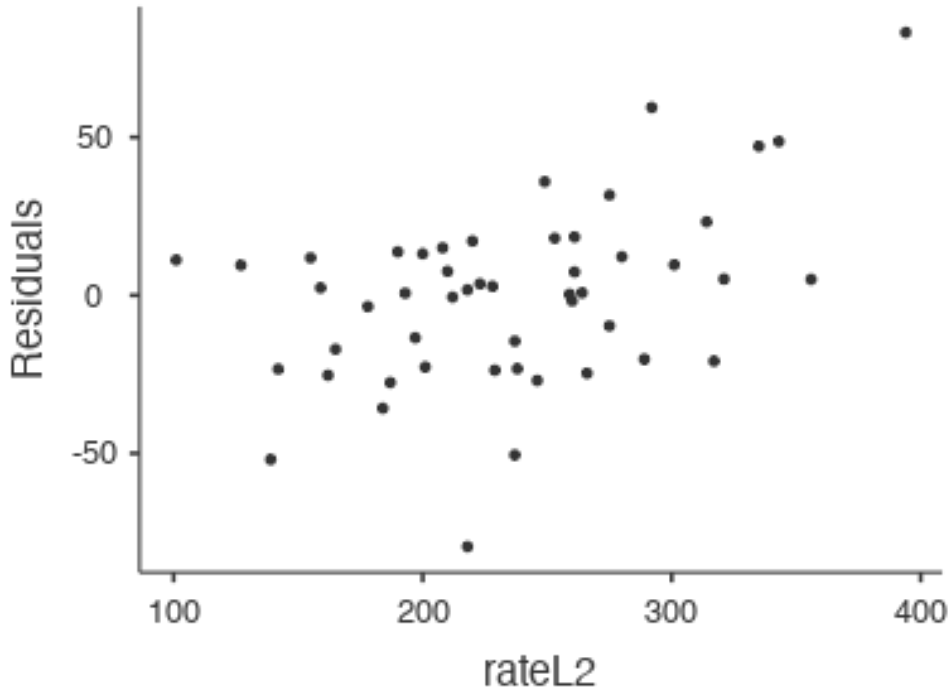

Figure 19. Residual plot for L2 looking time per word

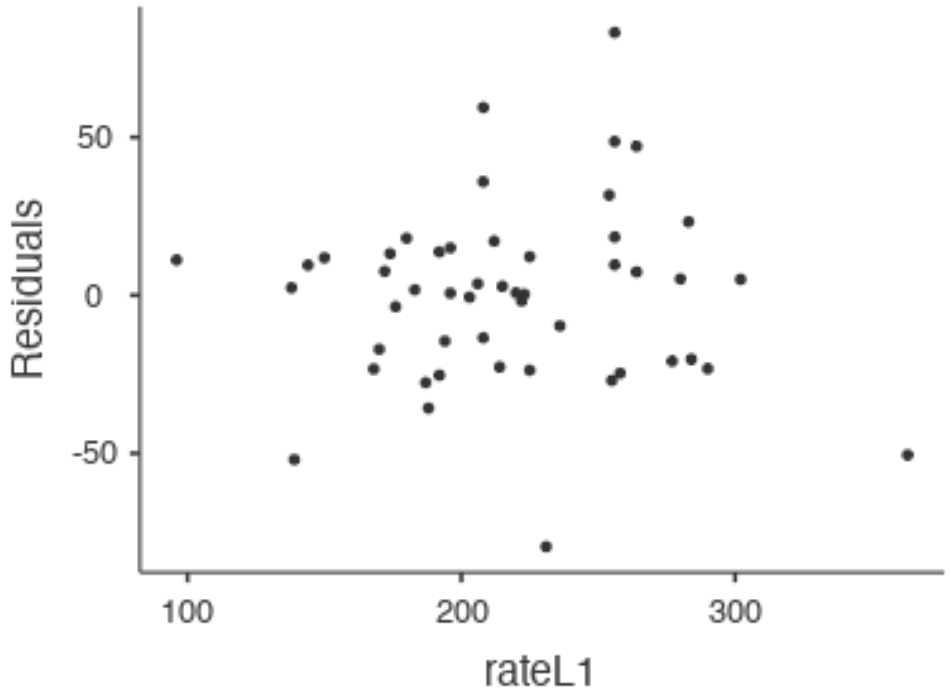

Figure 20. Residual plot for L1 looking time per word

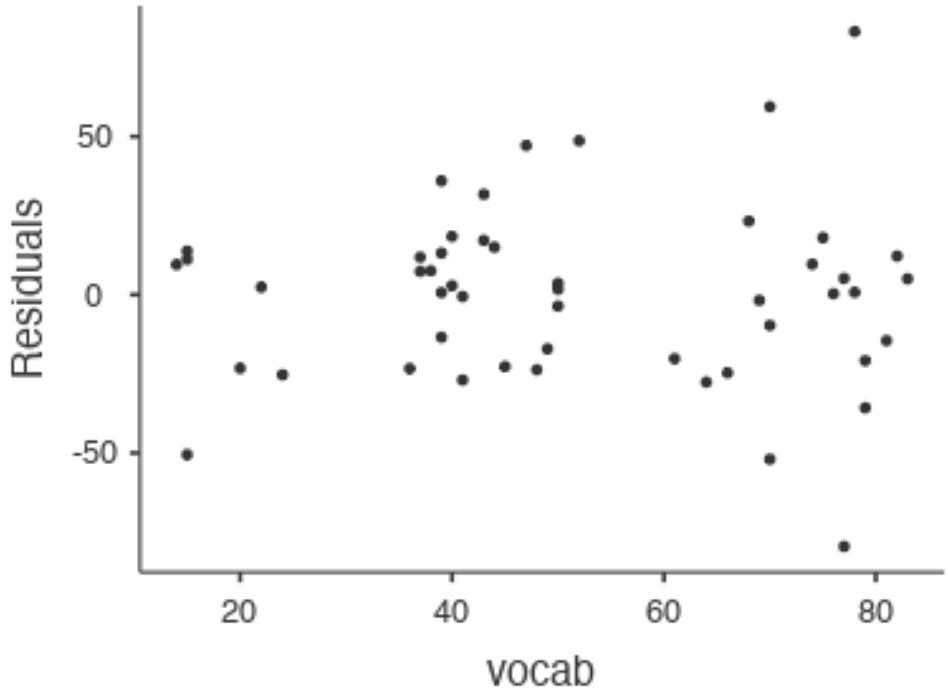

Figure 21. Residual plot for the vocabulary size test

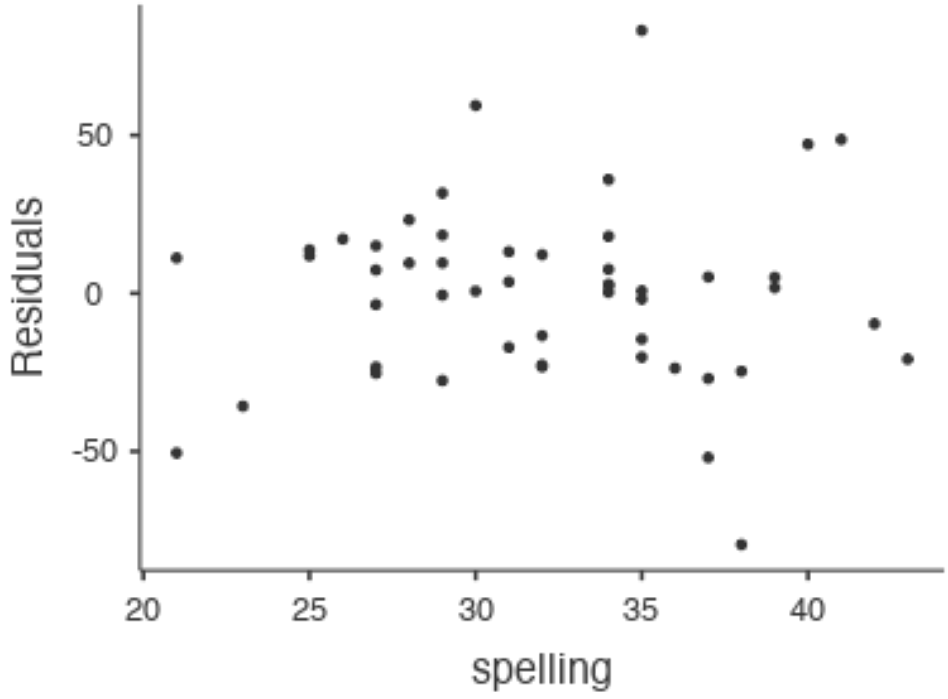

Figure 22. Residual plot for the spelling test

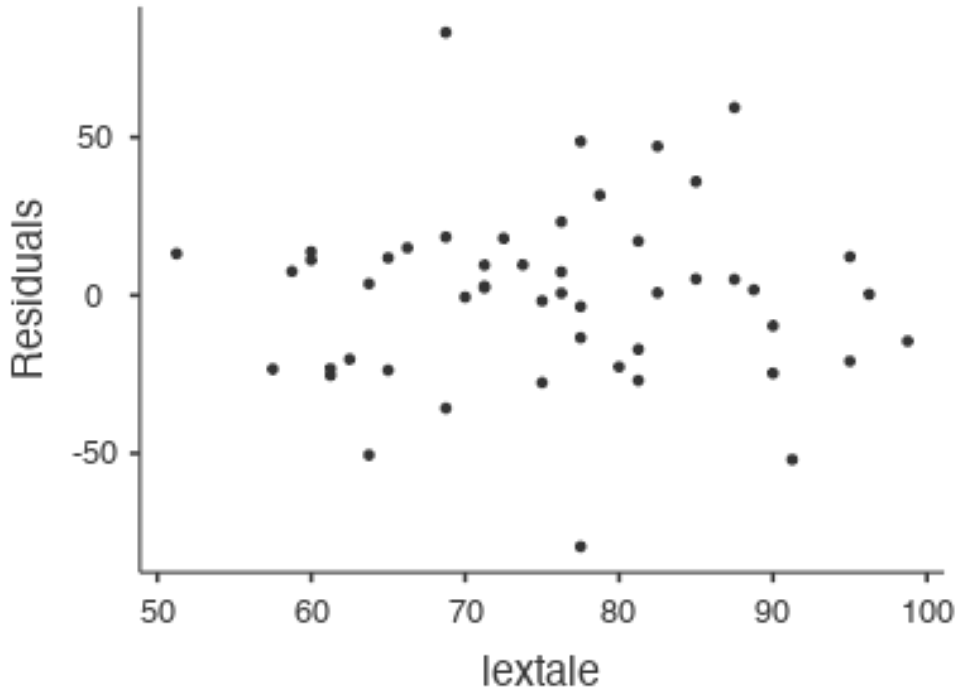

Figure 23. Residual plot for the lexical decision test

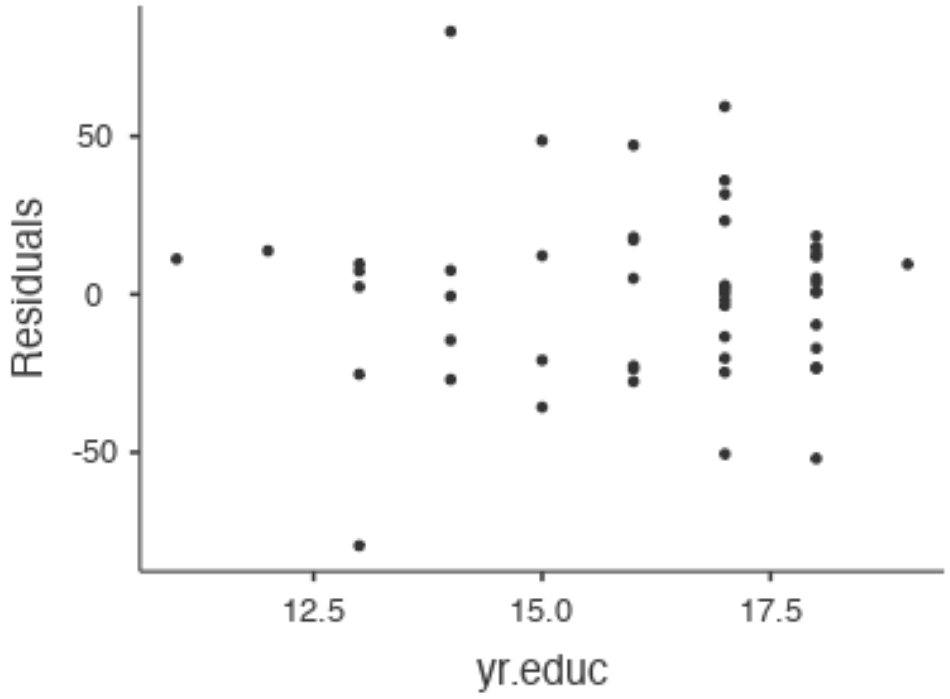

Figure 24. Residual plot for years of education