UNIVERSITY
OF OSLO

Master thesis

Identifying and Categorising the types of Reading Mistakes of children in early primary grades

November, 2023

Trinh Karen Alexandria Ramos Mai

SNE4390 Master’s Thesis

Credits: 40

Department of Education

Faculty of Educational Science

Identifying and Categorising the types of Reading Mistakes of children in early primary grades

Trinh Karen Alexandria Ramos Mai
November, 2023

Preface

First, I want to express my gratitude to the team in the teSTand project for allowing me to participate in their longitudinal study. It is an honour to work with you and be a part of your progress in creating a Norwegian standardised test, I thoroughly enjoyed collecting data and gaining experience through this project. I hope this project soon reaches its goal with only the greatest results. Second, I would like to thank my supervisors, Vasiliki Diamanti and Athanasios Protopapas, for all the guidance, patience and knowledge they have shared with me. I admire you for your passion in this field and it has inspired me to work even harder. Vasiliki, thank you for your kindness and trust in sharing this project with me. It was a pleasure to get to know you and to work with you.

I also want to thank my family and friends who have supported me every step of the way and encouraged me to do better. Thank you for taking an interest in my thesis and for listening to me talk about my thesis.

Table of Contents

Abstract ***5***

Preface ***3***

1 Introduction ***5***

1.2 Research question 5

1.3 Structure of the Thesis 5

2 Theoretical and Empirical Background ***6***

2.2 Phonological awareness 7

2.3 Alphabetic principle 9

2.4 Frith's three phases of reading 10

2.5 Ehri's five phases of word reading development 12

3 Models of reading development ***14***

3.1 Dual-route model of reading 14

3.2 Connectionist models of reading 15

3.3 The Bottom-up Model 17

3.4 The Top-down Model 17

3.5 The Interactive Model 17

4 Reading errors ***18***

5 Reading development across European languages ***20***

5.1 Norwegian orthography 23

5 Method ***29***

5.1 Aim of the study 29

5.2 Participants 29

4.2 Materials 29

4.3 Design 29

4.4 Procedure 30

4.5 Reading test 30

4.6 Reliability 31

4.7 Validity 32

4.8 Ethical implications 34

7 Discussion ***55***

7 References ***59***

Abstract

This study aims to identify and categorise the types of reading mistakes that children in Grade 1 to 3 make when they read a list of words and nonwords. Participants were presented with two lists of real words and two lists of nonwords, and were asked to read the words as quickly and as accurately as they could in 45 seconds. Their mistakes were scored, and analysed to be categorised. This study concludes with eight categories: Vowel length confusion, including vowel length shorten and length, consonant substitution, vowel substitution, addition, omission, syllabification, lexicalisation and other reading mistakes.

1 Introduction

1.1 Background and Rationale

An in-depth analysis of reading mistakes in early primary grades reveals distinct categories of reading errors, providing valuable insights for educators to tailor effective reading instructions and support young learners in their literacy development.

1.2 Research question

Based on previous research, this present study intends to focus on mispronunciations of letters and letter combinations within words that were misread in a reading task. This leads to the research question:

"Identifying and Categorising the types of Reading Mistakes of children in early primary grades"

1.3 Structure of the Thesis

Chapter 1 is a short introduction to what this thesis intends to examine based on the research question.

Chapter 2 is about reading development, including phonological awareness, alphabetic principle and the different theories of reading development.

Chapter 3 introduces the different types of reading models and how they demonstrate reading acquisition.
Chapter 4 introduces the different types of reading mistakes.
Chapter 5 is about the orthography of the different European languages, including Norwegian.
Chapter 6 shows the methodological approach that was chosen in this study with details on how the study was set up.
Chapter 7 shows the results gathered from the data and the analyses.
Chapter 8 discusses the findings from the study, limitation and future studies.

2 Theoretical and Empirical Background

Reading proficiency is fundamental for a child's academic success and overall development. The act of reading involves various abilities, strategies and knowledge. In the early stages of primary education, young learners undergo a crucial phase in their literacy journey where they acquire essential reading abilities. Reading mistakes are inevitable during this developmental period and these mistakes can serve as an indication of their reading progress or reveal potential problems in their reading development. Speaking often develops more naturally due to exposure to social environment, whereas reading is a skill that must be actively taught and learned through instructions and imitations. It is suggested that in order to acquire literacy skills, a child must be able to develop linguistic awareness. Acquiring both reading skills and writing skills go hand in hand as children are in the process of learning how to read the letters and words they are learning to write. These are therefore important skills to develop and are involved in two essential skills in learning how to read. The first is word decoding which relies on the child's alphabetic knowledge and phonemic awareness, while the other is comprehension of written language which refers to the understanding of the connection between speech sounds and letter combinations (Nergård-Nilssen, 2006).

Developing phonological awareness is crucial to developing reading skills as it combines the link between written words and spoken words. Phonological decoding is one of the most important skills which gives the reader the ability to read words aloud by combining letters and sounds. This can be assessed by evaluating nonword reading performances, as it challenges the reader to use acquired phonological knowledge to identify individual letters and their associated sounds.

Acquiring this ability allows the child to decode and reconstruct newer words during their first years of reading development (Hutzler, Ziegler, Perry, Wimmer & Zorzi, 2004). However, cross-linguistical evidence suggests that orthographic regularity greatly affects the child's ability to acquire phonological decoding skills due to the differences in syllabic complexity and orthographic depth. Children who learn how to read consistent orthographies seem to be able to decode unfamiliar words easier than children who read inconsistent orthographies like English in the early stages.

2.2 Phonological awareness

Phonological processing refers to how phonological information, as in sounds found in languages, is used in written and oral language (Wagner & Torgesen, 1987). It entails the skills of recognising and manipulating phonemes to form familiar and unfamiliar words (Seymour, 2005; Norton & Wolf, 2012) and also understanding that the sounds from spoken words can be broken down into smaller sound units (Goswami, 1999). Phonological awareness serves as a long-time predictor of reading development in several regular orthographies (Goswami, 1999) Lervåg, Bråten & Hulme, 2009). Studies of phonological processing have been divided into three research areas which are phonological awareness, phonological recoding in lexical access, and phonetic recoding in phonetic memory. Phonological awareness refers to the awareness and ability to break down speech sounds and access the phonology in a language. Well-developed phonological awareness is demonstrated through tasks such as counting the number of sounds in a word, reversing the sequence of sounds in a word, and blending isolated sounds to form a word (Wagner & Torgesen, 1987; Larsen, Hjetland & Schauber, 2022). Phonological recoding in lexical access is when written symbols are recoded into a system based on their sound representations. Phonological recoding can be assessed by tasks that require the reader to determine whether a letter string represents a real word or a pseudoword, or through tasks like rapid naming of colours, objects or numbers. While proficient readers can read high-frequency words by having lexical access that does not require phonological recoding, it is still important to have in the early stages of reading development and continues to be helpful when reading unfamiliar words In addition to recoding printed symbols into their representational sound-based system, the phonetic recoding enables the reader to maintain the letter-sound information in their working memory so it can be easily retained during processing. However, it has been suggested that skilled readers can comprehend words right after reading, which suggests that efficient phonetic coding is more important in the early stages of reading development. To master phonetic coding, the child must be able to decode written strings of letters, store sounds associated with the letters in their temporary lexical memory and be able to combine these sounds to create a word. As soon as readers have efficient phonetic decoding for storing letter sounds, they can apply more cognitive resources when combining sounds to form words (Wagner & Torgesen, 1987).

At the phonetic level, a single letter can be represented by several phonemes. Phonemes are known as speech sounds and are the smallest units in the spoken language (Wagner, 2008; Ehri et al., 2001). A combination of phonemes produces syllables and words which have different numbers of phonemes. Some words only have a singular phoneme, while most words have multiple. Short words like "a" or "oh" only have one phoneme, while other words like "go" (/gəʊ/) have two phonemes and "check" (/ʧɛk/) have three phonemes. The ability to manipulate these sounds means that one has developed phonemic awareness (Ehri et al., 2001). Phonemic awareness refers to the knowledge about the structure of a spoken word and the ability to recognise and categorise sounds (Harrison, 2004; Byrne & Fielding-Barnsleyy, 1989), and phonetic processing refers to the ability to recognise the connections between letters in written language and sounds in pronunciation (Ehri & Wilce, 1985). They are considered to be some of the main predictors regarding a child's reading proficiency and improve the development of reading acquisition when combined with grapheme-phoneme knowledge (Byrne & Fielding-Barnsley, 1989; Ehri et al., 2001). Graphemes are printed letters that represent a phoneme each, and there can either be one letter or a combination of letters representing one singular phoneme (Ehri et al., 2001). Readers who have acquired phonemic awareness have gained sufficient letter knowledge that they are able to differentiate phonemes in the pronunciations of whole words by isolating phonemes to pair them with correct graphemes. Those who have not mastered reading and writing are more likely to face problems during phonemic awareness tasks. Having phonological awareness also includes phonemic awareness, but also includes the ability to manipulate bigger units in spoken words such as syllables and rhymes (Ehri et al., 2001).

2.3 Alphabetic principle

The concept of the alphabetic principle is that written letters in a language represent sounds in the spoken language. The alphabetic principle applies to languages with alphabetical orthography such as English, which makes it possible to decode, but it does not mean that an English reader can automatically apply the alphabetic code when reading. Byrne (2014) demonstrates the understanding of the alphabetic principle by using an example of two English words “dog” and “den” which both consist of three letters, and based on the alphabetic principle, it also means that both words have three sounds. They also start with the same letter, which means both words start with the same pronunciation that is associated with the letter “d”. In short, the alphabetic principle states that a sound from a spoken language is usually represented by a letter in the written language (Byrne, 2014). Alphabetic coding therefore refers to the ability to link certain sounds to their represented letter, and it is one of the main indicators of reading comprehension (Sumbler, 1999; Nation & Cocksev, 2009). It is suggested that mastering how to decode words is a result of understanding the alphabetic principle. Readers who have learned to decode words that were not learned by sight such as irregular words or nonwords have most likely understood the alphabetic principle. However, a child may be unable to decode new words despite understanding the structure of “dog” and “den”, which does not necessarily imply that the reader was unable to understand the alphabetic principle (Byrne, 2014). Automatic alphabetic coding is often assessed by nonword reading tasks to observe whether the reader is able to decode new and unfamiliar words. If they do not perform well in nonword reading tasks, they are more likely to have difficulties decoding real words. They are instead more likely to rely on their visual memory by recognising sight words that they have already learned to decode before (Byrne, 2014).

Phonics is a strategy that involves implementing methods that expose beginners to the relationship between graphemes and phonemes, teaching them how to use written letters and sounds from spoken language to form words and read fluently (Amadi, 2019). There are two different approaches to this, one that is systematic and the other indirect. The systematic approach includes all elements of grapheme-phoneme correspondence in a specified order. It provides instructions that address short and long vowels, vowel and consonant diagraphs (combination of letters that represent one phoneme), and how to combine letters and sounds to form longer word subunits such as onsets and rimes (Amadi, 2019). An example of direct instruction is synthetic phonics which believes that it is needed for the reader to be aware of letter-sound relationships before they can start reading books and more. Beginners are taught how to decode new and unfamiliar words by sounding out printed letters and blending them. They learn how to combine letters in a written language with sounds in a spoken language in isolation, to form words and eventually learn to read more words that have similar letter-sound patterns to previously learned words. They learn to read sentences and stories as they develop their skills further. This approach excludes instruction of sight vocabulary (Amadi, 2019; Sumbler, 1999). Indirect approaches include analysing the relationship between letters and sounds by analysing whole words aided by context. The analytic approach is considered a mixed method as it focuses on decoding familiar words into smaller sound units like phonemes instead of learning the isolated pronunciations of words (Amadi, 2019). It refers to analysing a known word by breaking it down into smaller units, such as individual phonemes, grapheme-phoneme correspondences, or sound patterns. Usually, the reader should already have their sight-word vocabulary established as this approach teaches the letter-sound relationship through the context of known words. In general, children can naturally grasp the structure and rules of a spoken language simply by being exposed to an environment that actively uses the spoken language Sumbler, 1999; Byrne, 2014). This suggests that beginners do not necessarily need direct instruction or conscious effort to start learning the letter and sound relationships to read high-frequency and regular words but will need more letter-sound awareness for inconsistent words or orthographies.

2.4 Frith's three phases of reading

Uta Frith's (1958, as cited in Frith, 1986) theory of reading acquisition states that children undergo three developmental stages as they acquire the ability to read. These three phases are the logographic, alphabetic, and orthographic stages. In the initial stage, the logographic phase, young children heavily rely on visual cues such as shapes and colours of objects and symbols to recognise words. They may be able to identify logos, familiar words or their own name. At this stage, the child has not yet understood the relationship between letters and sounds, meaning they are unable to associate specific sounds with their corresponding letters or letter combinations, nor are they aware of the meaning of the letter order in words. Thus, the child primarily depends on colours, patterns, fonts and other visual features for word recognition (Simoncelli et al., 2004). As they progress, they move on to the next stage known as the alphabetic stage. During this stage, the child starts learning the alphabet and realises that combining singular letters to form words is possible. For example, they learn that the letters “c”, “a”, and “t” merged will form the word “cat”. They may start sounding out the individual letters to start forming the word. This ability is due to their development of phoneme awareness, which further helps them discover the relationship between letters and sounds and how these sounds combined can create spoken words. As they start developing their phoneme awareness, they may first start recognising the sound of the initial letter when faced with a printed word compared to the letters by the end of the word (Das, 2009; Simoncelli et al., 2004). Gaining an understanding of the grapheme-phoneme relationship therefore plays a crucial role in developing the ability to decode unfamiliar words or nonwords. Explicit phonics instructions should be provided to help the child read more accurately and independently. As their skills gradually become automatic, the reader can now process larger quantities of words without needing to sound the letters out to form words. When the reader can efficiently read words based on the letter arrangements and patterns, they move on to the final phase known as the orthographic stage. During this stage, the reader can now automatically connect sounds and letters to form them into words that are stored in their internal lexicon. Words in the internal lexicon are words that have been mentally processed due to repeated encounters and thus stored in their internal dictionary. These words are often referred to as sight words or vocabulary and are stored in their internal lexicon, thus making it easy for the reader to recognise a familiar word and also know its pronunciation and meaning (Harrison, 2004; Coltheart, 2006). The reader will soon manage to read lengthy texts and even comprehend unfamiliar words as they continue to improve these skills. It is therefore important that children are constantly exposed to diverse reading materials that can challenge and improve their reading proficiency (Firth, 1986).

2.5 Ehri's five phases of word reading development

Similar to Frith, Ehri (1998) introduced five phases of word acquisition: the pre-alphabetic phase, the partial-alphabetic phase, the full-alphabetic phase, the consolidated-alphabetic phase, and the automatic-alphabetic phase. Which phase a child is at reflects their current alphabetic knowledge. A child starts with the first phase, the pre-alphabetic, which resembles Frith's logographic phase where the reader has minimal alphabetic knowledge and has not developed the ability to map printed letters onto spoken sounds. They mainly depend on their memory and non-alphabetic cues, such as logos and pictures, as they are limited to sight word reading. They learn that words hold a meaning, but since they are unable to read them by print, they learn to associate pictures or signs with words. This phase is commonly referred to as the selective-cue stage (Ehri & Wilce, 1985) as their way of remembering how to read words is by their selected cues. Even if there is a letter change in a logo that a child is familiar with, they will still be able to recognise and read it. As long as the distinctive features remain visible, they will still remember which word it symbolises compared to when they are presented with printed letters only. This suggests that children rely on non-alphabetical visual cues more than actual letters as they rather read the surroundings instead of print. Children are also able to remember how to read words depending on the visual cues of the letters in a word, however, since they still have not connected the letters to sounds, they may not be able to read the word correctly. An example is how the word "look" might be easy to remember as the two "o" can resemble eyes, which are considered visual cues, but can be read as "see" if the letters are not associated with sounds in their memory (Ehri & McCormick, 1998). Children in kindergarten and first grade typically move on to the next stage, known as the partial-alphabetic phase, where children have developed an understanding of letters and have learned to associate some letters with their corresponding sounds. This is called partial-alphabetic because the reader is unable to divide the word up to all phonemes involved in the pronunciation. During this stage, readers often use the guessing strategy which combines their alphabetic knowledge with context cues when they are faced with unfamiliar words. As they now have learned about some connections between letters and sounds, they are able to recognise some words, but mostly by the first and last letters of the words which can confuse them if they try to read another word with the same initial and final letters. This makes it difficult for the reader to decode new and unfamiliar words (Ehri, 2005).

The full-alphabetic stage serves as an important phase before moving on to the next two phases. Children move on to this phase when they have learned sight words and have developed phonemic awareness meaning they have a better understanding of grapheme-phoneme correspondences. This gives them the ability to use spelling-sound techniques, or segment pronunciations, to decode new words and spell words better than children at the previous stage. The letter-sound relationships they easily recognise are words and sounds that they often encounter, and with practice, they can expand their sight vocabulary to more complex words which will help them read new words more efficiently as they have already learned how to read high-frequency words (Ehri & McCormick, 1998; Ehri, 2005). The fourth phase, the consolidated-alphabetic stage usually starts during the full-alphabetic phase but with a more developed phonemic awareness and bigger sight word memory (Ehri & McCormick, 1998). This means the letter-sound correspondences they have learned so far have developed into awareness of larger units, including onsets, rimes, syllables or affixes, which makes it helpful for the reader to read more advanced words with multiple syllables. This phase is important as the reader now requires less effort to make connections to store in their memory as the smaller units have now consolidated into bigger units (e.g., the word “interesting” consists of 10 grapheme-phonemes, and now turns into four syllables only) (Ehri, 2005). At the same time, the reader has developed a larger vocabulary including more complex words as they learn about morphemic suffixes, showing that they understand the meaning of -ed, -er, -ing, -est, -and, or -all, etc. by the end of a word. At first, beginners generalise morphemic suffixes by adding them at the end of any word, such as adding -ed to the word “soft” without awareness of grammar rules. Later they learn to use it for the correct grammatical category by using it only for verbs, and more specifically the right group of words, in this case, regular verbs (because of -ed) (Nunes, Bryant, Bindman, 1997). As their decoding skills also improve, they learn about hierarchal decoding and sequential decoding. Hierarchal decoding is the understanding of how a word's pronunciation changes when a grapheme is added or removed (e.g., in English, adding an -e at the end of a word: fin vs. fine; con vs. cone; or double consonants: cuter vs. cutter; diner vs. dinner). In Norwegian, the vowel pronunciation would differ between the words “søt” (cute) and “søtt” (sweet) as the double consonants shorten the vowel length. The final phase is called the automatic phase where the reader has now developed independent, accurate and efficient reading skills. Most words have now been stored in their mental lexicon and have developed their ways to decode unfamiliar or complex words. Teachers and adults can benefit from knowing about these five phases to provide appropriate instructions for their child's reading development (Ehri & McCormick, 1998).

3 Models of reading development

3.1 Dual-route model of reading

Visual word recognition helps the reader identify printed words that were previously stored in their mental lexicon. Proficient readers who have mastered this skill can easily convert printed letters into spoken words as they have managed to memorise pronunciations of familiar words and store them in their internal lexicon. This is known as the lexical procedure for reading aloud and it is theorised to involve two routes as outlined by the dual-route model (Coltheart, 2006). Through the mental lexicon, readers can access at least three types of information, such as word spelling, pronunciation, and meaning. The dual-route model introduces two different routes to access the internal lexicon. The first procedure is the lexical or direct route which has direct access to the mental lexicon, while the non-lexical/sub-lexical or indirect route relies more on using the grapheme-phoneme knowledge to identify and decode an unfamiliar word (Nergård-Nilssen, 2006). However, the model is more advanced than this and has therefore been elaborated in two ways. The first explanation introduces three types of lexicons: the orthographic lexicon which concerns the visual structure of a word, the phonological lexicon which concerns the pronunciations of a word, and the semantic system which refers to the meaning of a word. The use of lexical and non-lexical procedures depends on whether the reader encounters regular or irregular words. Regular words have a consistent relationship between letters and sounds, whereas irregular words do not follow the typical letter-sound relationship learned from reading regular words. Both lexical and non-lexical routes can be used to read regular words correctly, however, irregular words can only be read correctly by the lexical route as the non-lexical route would implement the letter-sound rules used on regular words onto the irregular words which would produce incorrect pronunciations (e.g., "said" would rhyme with "maid" and "have" with "cave"). Alternatively, it is possible to use the non-lexical procedure which has no access to the mental lexicon, and the reader will have to use their grapheme-phoneme knowledge to sound words out (Coltheart, 2006).

The dual-route model also introduces three effects: frequency effect, regularity effect, and lexicality effect. The frequency effect is generally considered a lexical effect as it retrieved the pronunciations of high-frequency words from the output lexicon quickly compared to low-frequency words. The regularity effect refers to how the letter-sound regularity and word frequency are important indicators of reading performance as it is suggested that reading low-frequency words is slower than reading high-frequency words. Words with irregular letter-sound correspondences were found to be more error-prone compared to words with regular letter-sound relationships (Ziegler, Perry & Coltheart, 2003). This effect is considered a sub-lexical effect because of how this route would implement regularisation of irregular words and result in mispronunciations of words irregular and exception words. Lastly, the lexicality effect refers to how real words, despite being regular or irregular, were read aloud faster than nonwords with the same number of syllables (Nergård-Nilssen, 2006). The dual-route model does not address the process of reading acquisition but rather views the reading progress to be quantitative. This means that even though most readers acquire the necessary components of the reading system, they vary in quantity rather than their qualitative qualities of reading abilities. If a child fails to master one of the lexicons at an age-appropriate level, their reading will be underdeveloped compared to their peers and their disadvantage will depend on the type of lexicon they are lacking. A limitation of this model is that it does not provide insights into how a child learns about grapheme-phoneme correspondences or how they store new words in their mental lexicon (Coltheart, 2006).

3.2 Connectionist models of reading

Connectionist models are computer-based models often used in research about reading and aim to present ideas about how reading is developed. The models are trained to recognise letters, letter strings, pronunciations, and word meanings. This provides the researchers with a valuable tool to use when testing their ideas and theories regarding reading development (Seidenberg, 2007). The networks consist of numerous artificial “neurons” which represent units, and “input units” refer to the printed letters of words, while “output units” refer to the pronunciation of words (Nergård-Nilssen, 2006). There are three main differences between the dual-route model and the connectionist model. First, the connectionist models assume that the representation of words is distributed, meaning that one word activates several units. Such units can respond to all phonemes in a word as all words contain a sound. In contrast, the dual-route model assumes that the words are locally represented in the reading system (Coltheart, 2006). Instead of distributing the smallest individual parts of the word, the locality refers to bigger items such as words as the model assumes that every word only has one single unit in the internal lexicon. However, other models have shown that both distributed and localised representations can work together within a model. The second main difference is their nature of processing. The connectionist model processes the information parallelly, meaning that when a reader encounters an unfamiliar word, all graphemes from that word will be processed simultaneously. The dual-route model processes serially where the sub-lexical procedure starts by converting graphemes into phonemes one by one from left to right. The last major difference between these approaches concerns learning. The models consist of networks of units that are connected. Once the first units are activated, it stimulates the other units. If a reader encounters an input as letter combinations, it activates another unit that contains phonetic characteristics. The spread of activation between units is determined by the connections between the units. Finding the appropriate set of weights, such as knowledge from experience, helps the model execute tasks more accurately (Seidenberg, 2007). If a connectionist model was to perform a reading-aloud task, it would need to be repeatedly exposed to correct spellings and pronunciations of words. When the model has mastered these words, the data will eventually perform more efficiently on the reading-aloud task. In contrast, the dual-route model is mainly used as a theoretical framework explaining the processes of learning development. Although the dual-route model uses knowledge that has gradually been collected by children, it is only designed to demonstrate how information processing develops in children (Coltheart, 2006).

3.3 The Bottom-up Model

The bottom-up model, also called the part-to-whole model, assumes that the reading process starts by decoding words into spoken sounds which later develops into higher levels of linguistic knowledge (Browne, 1998). The reader processes the visual input from the word, decoding the words and sentences based on which later progresses into the ability to identify words, structure sentences, and understand the meaning of the text (Browne, 1998; Amadi, 2019). This model is therefore mostly effective during early childhood as it prioritises phoneme awareness, such as learning about the relationship between letters and sounds. However, this model is at a disadvantage as the child matures since it mainly focuses on sight words rather than learning to read to understand the context. It has also been argued that children are able to learn about grapheme-phoneme correspondences independently without needing instruction beforehand (Browne, 1998; Baha, 2017). This would refer to the phonics approach that was described earlier, which highlights the importance of the relationship between letters and sounds of a language which they can later apply to reading and spelling (Amadi, 2019).

3.4 The Top-down Model

The top-down model or whole-to-part model highlights the importance of using pre-existing knowledge and contextual factors to understand the meaning of a text. In contrast to the bottom-up model, this approach perceives the reading process as constructing the meaning of a text based on prior knowledge. This approach requires the reader to have literacy knowledge, such as grammar and sentence structure to interpret the meaning of the text, and later use their knowledge about grapheme-phoneme correspondences to decode unfamiliar words. The main goal of this model is to teach children how to read for meaning as an act of communication through languages, which puts phoneme awareness as a lower priority (Browne, 1998; Baha, 2017).

3.5 The Interactive Model

The interactive model of reading instruction combines elements from both the bottom-up model and top-down model, which indicates that developing decoding skills and having prior literacy knowledge for comprehension is of equal importance in the reading process. This suggests that reading is an interactive process whereby the reader gains information from several sources to make sense of their reading. By developing grapheme-phoneme awareness, the child will learn how to decode and combine letters and sounds to form words. More reading experience leads to an automatic recognition of words, which increases the reader's capacity to shift focus on comprehending the text instead of decoding words one by one. This model acknowledges that reading is an interactive activity and stresses the importance of both decoding skills and reading comprehension in developing reading skills (Browne, 1998; Baha, 2017).

4 Reading errors

As stated, phonological decoding is essential in learning alphabetic languages as children learn to link letters to their corresponding sounds. If the reader is unlikely to master this skill, they will be prone to reading mistakes as their oral vocabulary knowledge might be underdeveloped. Reading a word differently from the printed text can be defined as a reading error as the reader fails to convey what the writer intended (Labov & Baker, 2010). Such mistakes can affect the overall meaning of the text and lead to mispronunciations. Analysing errors is important as it can shed light on what strategy the reader has chosen to use. Focusing on the partially correct factors within a reading mistake can reveal what kind of knowledge the reader has when identifying and pronouncing words. It can also provide valuable insights into potential skills they have not yet obtained (Weber, 1970; Leu Jr, 1982). It must be noted that reading errors must first be analysed at different levels of linguistic structure. An error in substituting a word for another suggests that the reader has failed to match sounds to corresponding letters and was unable to turn them into meaning (Weber, 1970).

There are two different ways of reading, silent reading and oral reading. Oral reading is important to evaluate during the first years of primary school as their reading skills can help identify reading mistakes. Oral reading is usually examined by how they decode printed letters, as the text includes their knowledge of phonetics, phonology and morphology (Labov & Baker, 2010; Erdem, 2017). There are different types of reading mistakes, specifically related to word recognition. The reader may be unfamiliar with the printed words which can lead to reading by guessing instead, or they have not yet learned the connection between letters and sounds in general or within the word. Many reading mistakes are due to mixing letters and words, changing letters within a word, adding or skipping words or letters, and inversions and repeating. Mistakes that include adding is when the reader includes an additional letter or word when reading, while skipping happens when the reader removes a sound, letter, or a whole word from their readings. Readers also tend to skip words they are unfamiliar with or skip prepositions (i.e., of, in, at) and conjunctions (i.e., for, and, but), or add affixes (i.e., -er, -ist, -ism) that were not in the text or are wrongly used with the word. These mistakes can be due to the reader reading too fast or lack of concentration (Erdem, 2017). A study by Basar and Batur (2018) examined the levels of oral reading by children around 5 years old compared to 6-year-olds who started attending primary school. Their results found that readers not in primary school made more oral reading mistakes, and their most common mistakes were either omitting a syllable or adding a syllable. This could be due to the younger readers not fully grasping the abstract rules of the language and not being fully aware of the relationship between printed letters and spoken language. It was also found that those who read words with closed syllables, meaning syllables with a short vowel and ends with a consonant, made more mistakes compared to reading open syllables (i.e., syllables with longer vowel sound and ends with a vowel) (Basar & Batur, 2018).

When a child lacks word recognition skills, they tend to make mistakes such as trying to repeat the words or part of the word until they succeed. They might repeat unfamiliar or longer words and syllabify the word (Erdem, 2017), as in dividing the words into smaller parts and repeating them until they successfully combine them into a new word. Inversions are one of the most common errors in beginners, especially in words with the letters “d” and “b” (Erdem, 2017), possibly because they are visually similar. In Norwegian, the word “do” turns into “bo” and they are both words with meaning. Another reading mistake involves substituting a word for another, which often happens between words that are phonologically or visually similar (Marinelli, Romani, McGowan, Giustizieri & Zoccolotti, 2023). This can also suggest that the reader is unable to connect the right letters and sounds and is not aware of how the meaning of the word fits with the text which causes misinterpretations (Weber, 1970). Sounding-out behaviour is also considered a common reading mistake but is also considered as behaviour as the child gradually corrects themselves until they successfully read the word right. This behaviour also shows that the child heavily relies on their phonological and non-lexical route, meaning that they use their knowledge about the letter and sound connection by decoding an unfamiliar word (Marinelli et al., 2023; Nergård-Nilssen, 2006).

5 Reading development across European languages

Spoken languages are visually represented by orthographic systems despite their differences, such as the alphabetic or other character-based systems and Braille. Young children learn to recognise sounds and sound combinations based on the languages that are mostly spoken around them or to them. As they mature, they develop an understanding of the sound structures of words (Goswami, 2008). Theoretically, it is assumed that if the graphemes correspond to the same number of phonemes, the easier it is for beginners to learn the orthography of a language. A wider gap between the number of phonemes and graphemes indicates a more complex relationship which makes it difficult for the reader to access the grasp the text through a phonemic approach and will therefore have to use other reading strategies that were mentioned above (Hagtvet, Helland & Lyster, 2006).

Many studies have examined the differences in reading fluency between children who have learned to read English orthography and children who can read other European orthographies. Seymour, Aro and Erskine (2003) categorised the European languages into two dimensions, those of “syllabic complexity” and “orthographic depth”. Syllabic complexity refers to the complexity of syllables as it measures how challenging it is to determine the boundaries of syllables in words (Adsett & Marchand, 2010), such as words with consonant-vowel syllables with a few initial or final consonant clusters. This includes Romance languages such as Italian and Spanish. The orthographic depth refers to the degree of consistency between letters and sounds, with more complex consonant clusters with closed consonant-vowel-consonant syllables.

This includes Germanic languages such as German and English (Seymour et al., 2003). Comparing orthographic depths shows the differences between the complexity and consistency in the relationship between graphemes and phonemes in different languages. Shallow or transparent orthographies, such as Italian, Spanish, Finnish and Turkish, have grapheme-phoneme correspondences that are more consistent than the English orthography which is a deep or opaque orthography (Nilssen, 2005; Seymour, 2005). Consistent orthographies have a one-to-one mapping between letters and usually have fewer exception rules (Perea & Estéves, 2008), whereas sounds while inconsistent orthographies have irregular words, multi-letter graphemes and context-dependent rules (Seymour et al., 2003). The English orthography is considered inconsistent due to the many possible pronunciations of the orthography to phonology which depends on spelling, grammar, or context (Nilssen, 2005; Seymour, 2005; Lervåg et al., 2009). Germanic languages, such as German and English, have a complex syllable structure that consists of closed CVC (consonant-vowel-consonant) with consonant clusters, while Romance languages, such as Italian and Spanish, have a simpler syllabic structure with fewer initial and consonant clusters and more open CV syllables (Seymour, 2005).

A cross-language comparison study conducted by Goswami, Ziegler and Scheiner (2003) introduced the idea of the flexible-unit-size hypothesis which refers to how the development of two reading strategies happens simultaneously. The first strategy is when the reader uses representations of small-sized phonemes that can be mapped onto single graphemes which reduces the complexity of orthographic clusters. The second strategy is to develop orthographic units that are bigger than the small-sized phonemes that can be represented by phonological rhymes which helps reduce the inconsistency of grapheme-phoneme in English words. The development of both strategies is called the flexible-unit-size hypothesis. The researchers hypothesised that children who learned to read in English were more likely to use both strategies compared to German children due to the inconsistency between the two orthographies. The English group and German group consisted of three groups of 24 seven- to nine-year-olds each and were asked to read two types of nonwords. The nonwords were categorised into two groups, blocked condition and mixed condition. In the blocked condition, large-unit nonwords and small-unit nonwords were separated into individual lists, while in the mixed condition, both large-unit and small-unit nonwords were presented in the same lists. Half of the children in each group received the first type of nonwords which consisted of large-unit nonwords that were orthographically similar to familiar real words (in English: dake, bicket, dactory; in German: dot, lenster, laramel). Readers could use rhyme units from the real words in their mental lexicon or use their knowledge about grapheme-phoneme relationships. The second type consisted of small-unit nonwords which were phonologically similar to real words, but not orthographically (in English: daik, bikket, dacktori; in German: dodt, länster, larramäll). The results suggested that English readers had to switch between the two strategies in the mixed condition (49% monosyllables were read correctly) compared to the blocked condition (67% monosyllables were read correctly), whereas German readers did no systematic switching and read 89% of the monosyllables correctly in the blocked list and 90% correctly read in the mixed list. English readers had better results when reading nonwords in the blocked condition compared to German readers where the blocked representations did not affect their used strategies and did not score better in reading large-unit nonwords. This suggests that German readers learn how to decode words by using the grapheme-phoneme correspondence which also includes decoding of larger grain-sized words. At the same time, the results can indicate reading the blocked list helps decoding both large-unit nonwords and small-unit nonwords which the English readers were most successful in using. Overall, English readers seem to be more flexible in how they switch between strategies, but it takes them a longer time to develop proper grapheme-phoneme correspondences that would increase their accuracy in reading small-unit nonwords and reading in general (Goswami et al., 2003).

Other cross-linguistic literatures have mostly compared the English language and another language with a regular orthography. Findings suggest that regularity does play an important role in reading development. It was also reported that children who learn how to read regular orthographies scored better in reading nonwords, high-frequency words and compound words compared to English readers with dyslexia. Despite being behind in learning transparent orthographies compared to their peers with no dyslexia, they eventually acquired the ability to decode new letter clusters, meanwhile, English readers continued to have difficulties in decoding nonwords in their adulthood as well (Nergård-Nilssen, 2005). A recent study by Marinelli, Romani, McGowan, Guistizieri and Zoccolotti (2023) examined the accuracy of reading performance of English and Italian readers in primary grades. The errors were classified based on Hendrik and Kolk's (1997) previous categorisations of reading errors: sounding-out behaviour and word-substitution errors. Sounding-out behaviour is when the reader tries to pronounce the target word gradually until they succeed, which mostly relies on the sub-lexical route using grapheme-phoneme correspondences. Word-substitution happens when the child reads another word that is visually or phonologically similar to the target word. Marinelli et al. (2023) added other categories for error analysis, such as "fragment" where only the first part of the word is read, "syllabication" where they decode and read the word in smaller parts, "regularisation" to compare common errors with opaque orthographies, and "lexicalisation" errors where a reader reads a nonword as a real word. Their results suggest that children who learn to read regular orthographies first rely on the sub-lexical route and eventually the lexical-route, while readers learning irregular orthographies start developing direct lexical access first. Sounding-out behaviour was mostly used by Italian children, showing slow progress that included several attempts until they either read correctly or incorrectly. However, English readers had more errors by substituting words due to irregular orthographies having more orthographical neighbours thus relying on lexicalisation (Marinelli et al., 2023).

5.1 Norwegian orthography

The Norwegian language, developed from the Old Norse, is a North Germanic language which belongs to the Indo-European languages. Norwegian orthography is described as semi-transparent due to its fairly systematic grapheme-phoneme correspondence. It is therefore possible for children at risk for reading development issues to learn the letter-sound relationship, but their development of reading speed and fluency can be negatively influenced by any inconsistency between the letter-sound correspondence and phonological variations, such as sound- and word length, and phonological- and orthographic complexity. Norway has two standard written languages: Bokmål ("book language") which is influenced by Danish, and Nynorsk ("new Norwegian") which is based on Norwegian rural dialects. The alphabet consists of 29 letters with around 40 phonemes. Twenty of these are consonant phonemes (B /be:/, B /se:/, D /de:/, F /ef/, G /ge:/, H /ho:/, J /je:/, K /ko:/, L /el/, M /em/, N /en/, P /pe:/, Q /ku:/, R /ær/, S /es/, T /te:/, V /ve:/, W /dobbelt ve:/, X /eks/, Z /set/), 9 vowels (A /a:/, E /e:, I /i:/, O /u:/, U /ʉ/, Y /y:/, Æ /æ:/, Ø /ø:/, Å /o:/) with either long or short pronunciations and five common diphthongs (AI /ai/, EI /æi/, AU /aʉ/, ØY /øy/, OY /oy/) [(Hagtvet, Helland & Lyster, 2006; Hillestad, Husby, Øvregaard & Robbins, n.d).

Due to some inconsistencies in the grapheme-phoneme correspondences, the Norwegian orthography is considered to be in between shallow and deep orthographies. There are several reasons as to why this orthography is considered complex. There are several consonant clusters found in many Norwegian words (*skjema* /'ʃeːmɑ/ (form); *nifts* /nɪfs/ (scary)). Usually, consonant clusters are stressed in the initial part of a multi-syllabic word, such as *skole* (school) and *merkelig* (strange) (Larsen et al., 2022). This is described as having a complex syllabic structure due to closed CVC syllables, including consonant clusters at the beginning and the end of a word (Seymour et al., 2003). There are also instances where combined letters are pronounced as one phoneme, such as *jeg*, *meg, deg* and *seg* are all pronounced with an "-ei" /æi/ after the initial letter and in the interrogatives, which is also found in the English language, in this case, the letters "hv" in Norwegian and "wh" in English: *hva* (what), *hvem* (who), *hvor* (where). Some letters can be pronounced in different ways, such as the letter "e", which can be pronounced as [i] in the word *de* (they), [e] in *det* (it/that), in *dem* [e] (them) and [æ] in the word *der* (there) (Nergård-Nilssen, 2005). The consonants *c, q, w, x* and *z* are regarded as foreign and are often used in loan words (e.g., taxi, chille) or are mostly replaced (e.g., zebra becomes *sebra*; October becomes *oktober*) (Larsen et al., 2022). Silent letters also exist in certain Norwegian words, such as the d in *land* and *blid*, or h in *hjelp* (help) and *hvit* (white) (Hagtvet et al., 2006; Larsen et al., 2022). The Norwegian orthography also demonstrates differences in vowel length depending on the subsequent consonants. If the following is a single consonant, the vowel sound is longer, whereas the following double consonants would make the vowel sound shorter (e.g., *tak* (roof); *takk* (thanks)) (Hagtvet et al., 2006; Larsen et al., 2022).

The inability to understand the vowel length rules in Norwegian is often an implication that a child is having trouble with their reading development or is prone to develop a reading disorder. Other than knowledge about vowel length, the study by Nergård-Nilssen (2006) introduces four other markers that are included in the reading development of Norwegian orthography based on their performances. The regularity effects refer to better performance in reading regular words compared to irregular words; frequency effects show advanced performance in recognising high-frequency words compared to low-frequency words; lexicality effects concern the reader's performance between real word and nonword reading; and granularity effects refer to the reader's strategy in small or large grain size reading depending on the consistency of the language (Nergård-Nilssen, 2006).

Nergård-Nilssen (2005) conducted a study following four Norwegian students with dyslexia from the age of two to eight and compared them with 23 ten-year-old children with no reading development issues. Their reading was assessed through several tasks: word reading, nonword decoding and text reading. In this study, the relevant tests were the word identification or naming task and nonword reading speed and accuracy. For the naming task, the child was presented with four individual lists of words they were asked to read aloud within 40 seconds. The words were monosyllabic with CV, VC, CVC and CVCC formats and bisyllabic words with CVCV, CCVCV and CCCVC consonant clusters. The lists consisted of regular high-frequency and low-frequency words, and the remaining lists consisted of high-frequency and low-frequency irregular words which included consonant clusters at the beginning and end of the word, and pronunciations of vowels were determined by the orthographic rules. The child was scored on how many words were correctly read in each list. For the nonword reading test, the child was presented with three lists of nonwords which they also had to read aloud within 40 seconds and were scored by how many words were read correctly. All subtests also included monosyllabic nonwords (CV, VC, CVC and CVCC formats) and bisyllabic nonwords (CVCV and CCVCV formats). The first subtest was to evaluate the reader's grapheme-phoneme skills in smaller units and was presented with nonwords with unusual syllables and initial and rime letters. The second subtest was the opposite consisting of nonwords with usual syllables, onsets and rimes that could be read as regular words, which examines the reader's ability to decode phonemes at a larger unit size. The last subtest tests the child's knowledge of orthographic rules, which was also made of nonwords with common syllables, onset and rime, but they could be read resembling real irregular words. They were scored by how many words were accurately decoded and read from each list. All these tests followed the Norwegian orthographic rules. Results of nonword reading of children with dyslexia showed that their accuracy score was within the range for non-dyslexic children. However, their performance suggests that they did not do as well in phoneme length awareness and phoneme quality awareness, meaning that they most likely lack awareness of the relationship between sounds and printed words. Inconsistency of vowel phonemes depends on the surrounding letters and the results showed that these children did not read them accurately. Similar findings were found regarding consonant clusters, suggesting that the children were unable to automatise the correspondences between letters and sounds (Nergård-Nilssen, 2005).

6 Nonword reading

As previously stated, phonological decoding is often measured by nonword reading fluency. This distinguishes poor readers from skilled ones as they tend to rely on context clues or initial consonant clues to predict the rest of the target word. Nonword reading is an assessment that requires the reader to break an unfamiliar word into smaller units and refer them to their associated sound in order to correctly pronounce a new word. This task is therefore effective at detecting if a reader is able to use their lexical knowledge to decode new and unfamiliar words. Fluent reading does not only depend on the ability to phonologically decode a new word but also on comprehension speed and reading accuracy. Rapid Automatised Naming (RAN) tasks are often used to assess automaticity as the tests require readers to rapidly name stimuli, such as pictures, letters, or words, that are presented to them. The subtest of reading letters or words is also useful when measuring how quickly a reader can retrieve phonological codes that are associated with graphemes (Fletcher, 2019). For nonword reading, TOWRE is commonly used to assess both accuracy and reading speed as the task is a nonword reading task. As decoding becomes automatic, more of the cognitive effort can be dedicated to processing the meaning of text (Wagner, 2008).

6.1 Nonword reading across orthographies

Bjaalid and Lundberg (1996) indicated that it was easy to create nonwords that could easily be decoded as real words in Norwegian due to its semi-transparent orthography. This was done by omitting silent letters from real words to form nonwords that would phonologically be similar. Their study examined 147 third graders across Norway who spoke Norwegian as their mother tongue. One of the four tests provided included the Phonological Choice Test (PCT) which included forty pairs of nonwords where in each pair, one of the nonwords was phonologically identical to a real word, while the other nonword had the possibility of sounding like a real word but was not phonologically akin to a known word. The results suggested that the reader's phonological ability was a strong factor in identifying poor readers, which is theorised to be important during the early stages of reading development (Bjaalid & Lundberg, 1996).

As previously found in the English language, vowels might be determined by consonants preceding or following. In Norwegian, the pronunciation of vowels depends on the consonants following. Double consonants after a vowel shorten the sound, while a single consonant after a vowel will lengthen the sound. This basic knowledge in Norwegian reading can affect how children initially read nonwords, and they are more likely to base their pronunciations on previously known words. If the nonwords are read incorrectly, they will have to use the mispronunciation correction process by breaking the phonemes down to adjust their mispronunciations which may not match words from their oral vocabulary (Murray, 2018).

The lexicon decision task asks the reader to decide whether a printed letter combination is a real word or a nonword. Usually, a pair of letters that normally would not be used in a real word together is enough to inform the reader that it is a nonword. Yet if the nonword resembles a real word, the reader will have to consult their mental lexicon which consists of known words the reader has accumulated over time. The ability to find the real word in the mental dictionary is defined as having visual word recognition. Performing this task can therefore help researchers examine how readers recognise words while reading. The simplest assumption is that the reader will have to confirm whether the presented printed word is a match to a word in their mental lexicon. However, this would take too long as expert readers would have more than 20 000 words stored in their internal vocabulary (Coltheart, 2006).

Landerl (2000) questioned whether different reading instructions mattered in nonword reading between contrasting orthographies such as English and German. Participants were divided in three groups where the first English group received mixed instructions that combined phonics approach and whole-word reading methods, while the second English group solely received phonics teaching approach which can be similar to reading methods used by German speaking countries. The last group consists of German readers who uses the phonics approach due to their consistent orthography. One of the scoring methods used was based on Wimmer and Goswami (1994) were they marked nonword pronunciations as correct if it was based on an analogy of a real word. The nonword "nour" would therefore be scored as correct if the pronunciation rhymed with the real words "our", "tour" and "four", or pronunciations that rhymed with "five" or "give" for the nonword "twive". However, the English phonics group seemed to be in disadvantage by this type of scoring, but reading the German nonwords would sound identical whether the reader rhymed with real words or based on the grapheme-phoneme rules due to the consistent orthography. Due to this disadvantage, the researcher decided to add a more lenient scoring method where any sound-to-spelling reading similar to an existing English word were consider correct. The nonwords "sen" could be read with a long *e* (/si:n/) based on the English alphabetic pronunciation of the letter *e* and pronounced in the nonword "twive", where the correct pronunciation would have a silent *e* if marked by Wimmer and Goswami's (1994) scoring method. The results showed that English readers who received phonics instruction had more nonword reading errors compared to the German group, but would have a lower error rate if scored leniently. Overall, the English group that received phonics instruction was more accurate in nonword reading compared to English readers who received mixed instructions. This suggest that a reading instruction similar to what consistent orthographic readers use may have a positive effect on nonword reading, even for an inconsistent orthography like English (Landerl, 2000).

5 Method

5.1 Aim of the study

The first years of primary school are crucial for reading development in general and are important for correcting reading errors. Their reading mistakes at this stage can inform teachers and parents what kind of instruction and guidance they might need in the future.
Oral reading is important to develop during primary grades.

5.2 Participants

The current study consisted of 95 participants, out of which 17 (18%) were pupils in Grade 1, 25 (26%) in Grade 2, and 53 (56%) in Grade 3. The participants were recruited from 10 different schools across Oslo, Viken and Vestland regions with half of the schools being situated in urban areas and the other half in suburban areas. The participants had to be efficient in spoken Norwegian. Non-native speakers of Norwegian had to have attended kindergarten for three consecutive years in order to be included in the sample. Participants joined voluntarily with written consent from their parents. Data in this study was gathered between 2021 and 2022. All personal information about the participants was anonymised and saved in a secure digital database (TSD).

4.2 Materials

An iPad application was developed for the teSTand project's purpose of digitalization and standardisation of tests in Norwegian. The application included several tasks of phonological awareness, morphological awareness, numeracy, RAN and reading. The app demonstrated examples and practice rounds before the administration of the actual tasks. For the present thesis, only the data from the reading test was used.

4.3 Design

The study aims to identify and categorise the types of reading mistakes found in Grade 1, Grade 2 and Grade 3 when reading a list of real words and nonwords. The purpose of the task is to assess reading in Norwegian orthography by presenting children with two lists of words and two lists of nonwords that they were asked to read within a 45-second timeframe. This is an experimental quantitative study. The independent variable, Grade, consists of three levels which are Grade 1, Grade 2 and Grade 3. The dependent variables were the mean scores of correct answers and percentages of errors from identified categories of reading mistakes. To conduct the study, the examiner and the child were provided a quiet room at school as the child had to be recorded while taking the test. Younger children were offered two sessions to complete the reading tests, but all participants completed them in one session each.

4.4 Procedure

A trained examiner and a child were provided private rooms by the school to take the tests confidentially and without distractions. The child was registered before starting the test which was carried out digitally and on paper. Before the test, the child was presented with five example words for each list to read out loud and would receive feedback if they read a word incorrectly. For the test, the child was given two printed lists of words and two printed lists of nonwords. They were asked to read as many as they could within 45 seconds for each list. The examiner had the lists displayed on the iPad to mark which words were read incorrectly while the iPad recorded the child's reading. If the child stopped or had a long pause at a certain word, they were urged to move on to the next word. The examiner could in no way aid or give hints about the correct pronunciations during the procedure. The purpose of the recordings is so the examinator can review and mark the answers appropriately, take notes of mispronunciations and time how long it takes the child to recognise or read a word.

4.5 Reading test

All three grades were presented with two lists of real words (see Appendix) and two lists of nonwords (see Appendix). Each list of real words consists of 101 words, and each list of nonwords consists of 63 words. Participants were asked to read aloud all four lists of words from top to bottom as accurately and as quickly as they could within 45 seconds per list. The test is inspired by the Test of Word Reading Efficiency (TOWRE) (Torgesen, Wagner & Rashotte, 1999). The assessment consists of two parts, starting with the practice round with eight words for the child to familiarise themselves with the task. The main part includes lists of 101 words each, presented vertically on an A4-sized page. The words were chosen based on specific criteria such as the number of syllables, syllabic structure, phonemic complexity, orthographic complexity, and frequency of occurrence in the Norwegian subtitles corpus. The lists of words introduce very high-frequency words early in the list and progressively less frequent words further down the list, arranged in an increasing order of syllables. The words were chosen with the help of an online tool called the Norwegian Orthographic Analyzer (NOA) which has a word database composed of subtitles from Norwegian media. This tool is used for finding words with pre-specified criteria and/or for extracting information about given words, such as word frequency and other orthographic, grammatical and statistical properties of Norwegian words (Norwegian Orthographic Analyzer, 2023). Additionally, participants were also presented with two lists of nonwords, each list consisting of 63 items. The requirements for creating the nonwords were based on the same criteria as the real words, which are the number of syllables, syllabic structure, phonemic complexity and orthographic complexity. The pseudowords were also arranged in increasing order of syllables, but with fewer items and a maximum length of three and four syllables, while the real words reached six syllables. The items were presented in three rows with 21 words in each column and were read from top to bottom and left to right. Participants were asked to read the lists of words and nonwords aloud as quickly and as accurately as they could within 45 seconds and were scored by how many words they read correctly.

4.6 Reliability

Reliability is measured to ensure that the probability of achieving similar results is possible by applying the same methods. There are different ways to measure whether a measurement is consistent or replicable. The test-retest reliability refers to whether the results will remain consistent over time by repeating the measurement afterwards. This means reusing the same test on the same group of people after a later time. The inter-rater reliability refers to the consistency across people, which questions whether observations and answers for the same questions happen to be similar across participants. The parallel forms reliability is to measure if the same answers are produced by different instruments, measuring the consistency across theoretically-equivalent measurements. Internal consistency reliability refers to the consistency of responses in multi-item scales, such as personality questionnaires, and whether the responses to different items are related or not (Navarro & Foxcroft, 2019).

4.7 Validity

While reliability measures the consistency of an assessment, validity tells us the accuracy of the measure and whether the results represent the aim of the study. Statistical analyses are essential in research, usually used to determine the correlation of hypothesised cause and effect. Random error and degree of covariation in the sample data are usually compared to figure out the covariation in the study. Most commonly, the probability level lies around 5% (alpha .05). If there is a need to prove covariation more, the researcher can lower the probability level to protect the study from error in covariation. Comparisons that lie at the probability score of .05 or below are considered to be significantly "true" while those with higher probability are considered "false". Statistical conclusion validity concerns whether the presumed covariation is plausible depending on the decided alpha level and acquired variances. Cook and Campbell (1979) pointed out some threats to statistical validity, such as low statistical power, violation of assumptions, fishing and error rate problems, low reliability and more. Low statistical power happens when the alpha level is too low and with a small sample size, as this increases the chance of a false-negative error (Type II error) and thus falsely accepting the null hypothesis. Assumptions of statistical tests are usually assessed to see if they meet the requirements for the results to be considered significant when there is a null hypothesis testing. It is therefore important to assess the assumption of normality and/or homogeneity when conducting statistical analyses to determine whether given results are valid regarding the study sample (Cook & Campbell, 1979). In this current study, assumption checks were conducted to test if the normality and homogeneity of variance were violated. Since ANOVA is a parametric test, which means it assumes data are normally distributed, it is important to assess the assumption of normality to decrease the chance of false positive results. If the assumption of normality is violated, other analyses are to be conducted. The Kruskal-Wallis test was conducted instead of a one-way ANOVA and Spearman's rho for correlation in this study.

Among different types of validity, internal validity and external validity are considered most important (Navarro & Foxcroft, 2019). Internal validity refers to the extent to which a method can accurately measure what it aims to measure. Achieving high validity means that the findings are representations of the world outside the study. Having high reliability is therefore important as it can indicate if a method is valid. However, reliability itself cannot measure whether findings are accurate representations of the real world. Internal validity focuses on the internal structure of the study. It assesses the causal relationships between variables without interference from outside factors, making sure that the changes in the dependent variable, as in the outcome, are due to the independent variable (i.e., the cause) (Cook & Campbell, 1979). To achieve high validity, researchers need to be aware of possible threats and minimise them as much as possible. There are several threats to internal validity such as history, maturation, testing effect, instrumentation, statistical regression, attrition, selection, experimental bias and more. Fortunately, it is possible to minimise these threats by manipulating the study. Randomisation of participants and/or assignments is often used to eliminate bias in the study. The use of control groups is also helpful in comparing groups that were and were not exposed to experimental conditions. In this case, the design of the study cannot accurately prove that a certain school grade performs better than others in reading real words and nonwords. This is due to variability in human behaviour and knowledge which can affect the child's reading skills.

On the other hand, external validity concerns the generalisability of results. It refers to what extent the results of a research study can be applied to the general public and other settings. If the results do not seem to apply to the general public or other target populations, it is likely lacking external validity. The external validity is threatened depending on the sample population choice, such as having a small sample size or a participant pool that is systematically different from the general population. One of the main threats to external validity is the interaction of causal relationships with units, meaning that the examined sample may influence the relationship between independent and dependent variables. Randomised selection of participants is commonly used in research to enhance the external validity of the study as well. This ensures that the chosen sample of participants is representative of the target population that is being examined (Ferguson, 2004; Findley, Kikuta & Denly, 2021). Cook and Campbell (1979) suggest that researchers look at the homogeneity of groups before the onset of the study and again towards the end. To avoid complications of the assumption of generalisability, researchers tend to enhance generalisability by determining the statistics of the representativeness of the sample. They might also only recruit participants who have fulfilled the requirements or have the characteristics of the targeted population (Ferguson, 2004).

Good construct validity refers to what extent the chosen method of measurement measures what it is intended to in terms of the theoretical construct (Navarro & Foxcroft, 2019). It helps find the appropriate types of measurement required to develop a study that examines the aim of the study. If the researcher chooses to use a questionnaire, they need to know whether that specific questionnaire assesses what they intend for it to measure. It is therefore important that the researcher bases their hypotheses and research design on relevant studies with similar goals (Ferguson, 2004). The target population in this study were students from 1st grade to 3rd grade enrolled in Norwegian schools. It was also required that they be fluent in the Norwegian language. The homogeneity and assumptions of normality were checked for each of the analyses conducted in this study. Correlation tests were therefore conducted to assess the relationships between the independent and dependent variables. More specifically, we examined the relationship of correct scores between the cumulative scores of Real Words and Nonwords (real words form A and form B, and nonwords form A and form B).

4.8 Ethical implications

The present study has been a part of a research project called teSTand, which is approved by the Data Protection Services at Sikt (formerly Norsk Senter for Forskningsdata or NSD). Schools, teachers and parents received an information sheet regarding the purpose of the study, and only children with signed consent from parents were eligible to participate. Children and parents were informed that all personal data will be pseudonymised and that they have the right to withdraw from the study at any time, and their personal information will be deleted after a withdrawn consent and/or by the end of the study. Data collectors received training from the teSTand research group and had a relevant background within Special Needs Education.

6 Results

Descriptives

Frequencies of grade

grade	Counts	% of Total	Cumulative %
G1	17	18 %	18 %
G2	25	26 %	44 %
G3	53	56 %	100 %

There are in total 95 participants in this study. Grade 1 consists of 17 children (18%), Grade 2 of 25 children (26%), and Grade 3 of 53 (56%). The independent variable is Grade, consisting of three levels: Grade 1, Grade 2 and Grade 3. The dependent variables for different analyses are mean scores and percentages of correct answers for each test (Real Word A, Real Word B, Nonword A and Nonword B) and cumulative scores for Word and Nonword sums, as well as the mean percentages of error mistakes in each category (Vowel Length Lengthen, Vowel Length Shorten, Consonant Substitution, Vowel Substitution, Syllabification, Addition, Omission, Lexicalisation, and Others). Unanswered words were also counted as errors.

6.1 Structure of the data

Assumptions for statistical tests were examined to determine which tests would be appropriate to conduct. While doing the analyses for this study, the assumptions of normality and homogeneity of variances were checked. Some of the variables violated the assumptions of normality and non-parametric tests were therefore conducted for all analyses despite the variables that did not violate the assumptions of normality. This is so results are directly comparable, thus conducting Spearman's rank correlation and Kruskal-Walling with Dwass-Steel-Critchlow-Fligner (DSCF) tests instead of parametric tests such as Pearson correlation coefficient and One-way ANOVA. First, correlation analyses were conducted to examine the relationship between grades and the mean scores and percentages of correct answers. The percentages of reading errors were examined by ANOVA analyses to see whether there were significant differences in error mistakes across grades. The statistical software Jamovi was used to conduct all the tests provided below. The figures demonstrate the relationships between variables found in these tests along with tables that show the results from each test and their *p*-value.

6.2 Analyses of correct answers

6.2.1 Correlations

In research, correlation analyses are often conducted to assess the statistical reliability and validity of a study. It measures whether there is a relationship between two variables. Since non-parametric tests are used in this study, Spearman's rho is used instead of Pearson correlation coefficient. While the Pearson's correlation coefficient measures the direction and significance of variables from assumed normal distribution, the Spearman's rho refers to the strength of the relationship between the ranks of the variables. This test is used to examine whether the variables covary, meaning that if one variable systematically increases, the other systematically decreases. The correlation coefficient ranges from 1 to -1 where 1 indicates positive correlation while -1 means negative correlation. If there is a positive relationship between two variables, both values increase.

Real Word A and Real Word B

Correlation Matrix

		RWa.OK	RWb.OK
RWa.OK	Spearman's rho	—	
	p-value	—	
RWb.OK	Spearman's rho	0.97 ***	—
	p-value	< .001	—

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

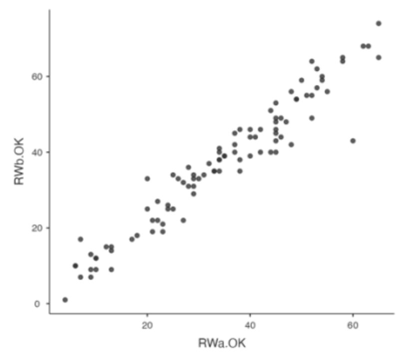

Figure 1. Scatterplot demonstrating a strong and positive relationship between mean scores of Real Word d A and Real Word B.

To examine the alternate-form reliability, Spearman's correlation was used to assess the correlation. Results show that there is a statistically significant relationship between the two tests ($\rho(93) = 0.97$, $p<.001$). A high correlation coefficient which is close to 1 indicates a strong and positive relationship between the variables. This can also be seen on Figure 3 which shows the values increasing.

Nonword A and Nonword B

Correlation Matrix

		NWa.OK	NWb.OK
NWa.OK	Spearman's rho	—	
	p-value	—	
NWb.OK	Spearman's rho	0.91 ***	—
	p-value	< .001	—

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

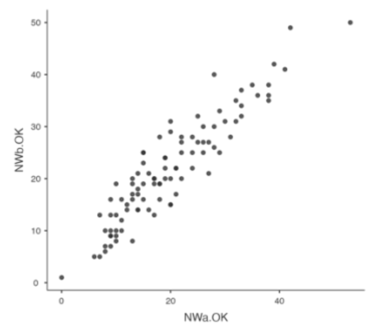

Figure 2. Scatterplot demonstrating a strong and positive relationship between mean scores of Nonword A and Nonword B.

As shown in the table, Spearman's rho shows that there is a significant correlation between the mean scores in Nonword A and Nonword B ($\rho(93) = 0.91$, $p<.001$). The correlation coefficient indicates a strong and positive relationship as it is close to 1, which is also demonstrated by the scatterplot in Figure 4. Including the previous test, both results show that there is a high reliability in this study.

6.2.2 ANOVA

Analysis of Variance (ANOVA) tests are used to compare the means of two or more groups. Here, the ANOVA tests are used to compare whether there is a difference in mean scores of Real Word and Nonword reading between Grade 1, Grade 2 and Grade 3 by analysing the levels of variance within the groups. The null hypothesis states that the data samples in all grades come from the same population with the same mean values. To make results comparable, the non-parametric one-way ANOVA test called the Kruskal-Wallis test was used for all analyses along with the Dwass-Steel-Critchlow-Fligner (DSCF) pairwise comparisons. The Kruskal-Wallis test examines the differences between groups with the data ranked from lowest to highest and whether the ranks are evenly distributed across the three grades. However, this test does not identify the differences in groups, which is why the DSCF test is used to compare the means of pairs across all grades.

Real Words

Kruskal-Wallis

	χ^2	df	p
RW.OK	34.69	2	< .001

Pairwise comparisons - RW.OK

		W	p
G1	G2	5.91	< .001
G1	G3	7.19	< .001
G2	G3	4.68	0.003

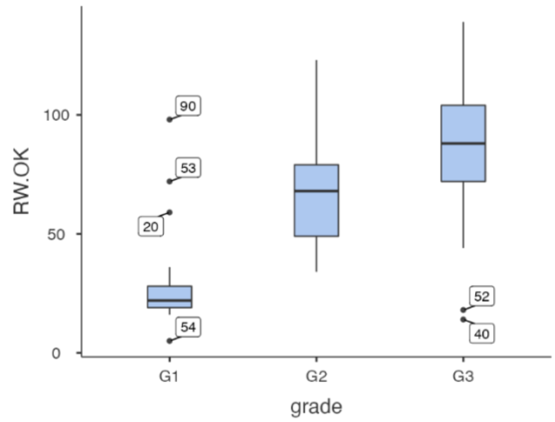

Figure 3. Box plot demonstrates the distribution of Real Word data across Grade 1, Grade2 and Grade 3.

The non-parametric ANOVA test, the Kruskal-Wallis test reports that there are statistically significant differences between the mean scores of all three grades ($\chi 2(2) = 34.69$, $p<.001$). Results from the DSCF test also show that there are significant differences between all grades: Grade 1 and Grade 2 ($W = 5.91$, $p<.001$), Grade 1 and Grade 3 ($W = 7.19$, $p<.001$), and Grade 2 and Grade 3 ($W = 4.68$, $p = .003$). The mean score of Grade 1 is 31.34, the mean score of Grade 2 is 66.64, and the mean score of Grade 3 is 86.19. The box plot in Figure 5 demonstrates that there is less variability in Grade 1 compared to the two other grades. Grade 3 seems to have some outliers as well. It also shows the ranges between grades where Grade 3 has a significantly larger range of mean scores compared to Grade 1.

Kruskal-Wallis

	χ^2	df	p
RW.OK.percent	14.33	2	< .001

Pairwise comparisons - RW.OK.percent

		W	p
G1	G2	3.88	0.017
G1	G3	5.11	< .001
G2	G3	1.82	0.401

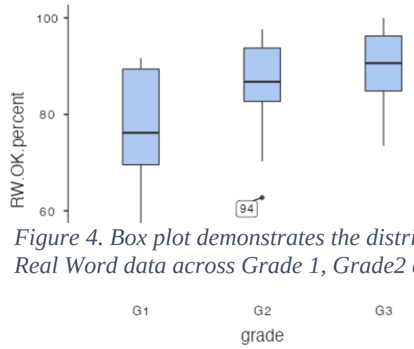

Figure 4. Box plot demonstrates the distribution of percentages in Real Word data across Grade 1, Grade2 and Grade 3.

Looking at the percentages of correct scores in Real Words, the Kruskal-Wallis test reports significant differences between the percentages of scores for all three grades ($\chi 2(2) = 14.33$, $p<.001$). There are significant differences between Grade 1 ($M = 77.47$) and Grade 2 ($M = 86.97$) ($W = 3.88$, $p = .017$) and Grade 1 and Grade 3 ($M = 88.66$) ($W = 5.11$, $p<.001$), but not Grade 2 and Grade 3 ($W = 1.82$, $p = .401$). The box plot shows a wider range in Grade 1 compared to the two other grades with outliers.

Nonwords

Kruskal-Wallis

	χ^2	df	p	ε^2
NW.OK	30.60	2	< .001	0.33

Pairwise comparisons - NW.OK

		W	p
G1	G2	5.84	< .001
G1	G3	6.94	< .001
G2	G3	3.84	0.018

Pairwise comparisons - NW.OK

		W	p

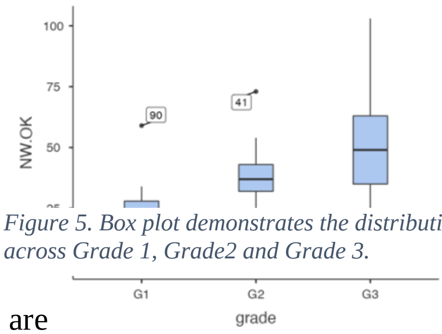

Figure 5. *Box plot demonstrates the distribution of Nonword data across Grade 1, Grade2 and Grade 3.*

The Kruskal-Wallis reports that there are statistically significant differences between all three grades ($\chi2(2) = 30.60$, $p<.001$). DSCF results also show that there is a significant difference between the pairs of all grades: Grade 1 ($M = 21.82$) and Grade 2 ($M = 38.00$) ($W = 5.84$, $p<.001$), Grade 1 and Grade 3 ($M = 49.87$) ($W = 6.94$, $p<.001$), and Grade 2 and Grade 3 ($W = 3.84$, $p = .018$). Figure 5 shows outliers in Grade 1 and Grade 2. It also shows a larger range of distribution of data in Grade 3 compared to Grade 1 and Grade 2 when reading nonwords.

Kruskal-Wallis

	χ^2	df	p
NW.OK.percent	1.82	2	0.402

Pairwise comparisons - NW.OK.percent

		W	p
G1	G2	1.20	0.675
G1	G3	1.71	0.446
G2	G3	1.04	0.744

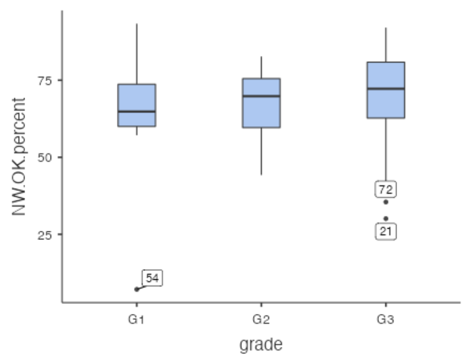

Figure 6. *Box plot demonstrates the distribution of percentages in Nonword data across Grade 1, Grade2 and Grade 3.*

Kruskal-Wallis test reports no significant differences in percentages of scores across all three grades ($\chi2(2) = 1.82$, $p = .402$). The DSCF test also shows no significant differences between the pairs: Grade 1 ($M = 65.28$) and Grade 2 ($M = 68.61$) ($W = 1.20$, $p = .675$), Grade 1 and Grade 3 ($M = 69.83$) ($W = 1.71$, $p = .446$), and Grade 2 and Grade 3 ($W = 1.04$, $p = .744$). The box plot on Figure 6 shows little variety across grades.

6.3 Analyses of error percentages

The errors were identified and divided into 8 categories: Vowel length confusion (VLC) with the subcategories Vowel Length Lengthen (VLL) and Vowel Length Shorten (VLS) as children did not pronounce the appropriate vowel phoneme length depending on the following consonants; Consonant Substitution (CS) and Vowel Substitution (VS) refers to incorrectly reading a word due to consonant or vowel shifting; Syllabification (SYB) refers to when children would break the alphabetic code, mostly seen in words with consonant clusters; Addition (ADD) and Omission (OMI) occurred when the child either added a phoneme or removed a phoneme in the given word; Lexicalisations (LEX) were found when a nonword was read as a real word with phonological similarities; Others (OTH) consist of errors that were not categorised. Unanswered words were also counted as errors.

6.3.1 Real Word errors

Descriptives

Descriptives

				Shapiro-Wilk	
	N	Mean	SD	W	p
RW.nErr	95	8.55	6.24	0.88	< .001
RW.VLL.percent	95	0.93	1.76	0.60	< .001
RW.VLS.percent	95	1.79	2.23	0.80	< .001
RW.CS.percent	95	3.36	4.00	0.79	< .001
RW.VS.percent	95	2.23	2.64	0.79	< .001
RW.SYB.percent	95	0.67	2.04	0.36	< .001
RW.ADD.percent	95	1.44	2.16	0.72	< .001
RW.OMI.percent	95	1.45	2.24	0.69	< .001
RW.OTH.percent	95	0.37	0.83	0.51	< .001
RW.UA.percent	95	1.53	4.55	0.37	< .001

Descriptives

	N	Mean	SD	Shapiro-Wilk W	Shapiro-Wilk p

Percentages of reading mistakes found in all grades reading the Real Word lists were calculated and found in all error categories concerning reading of real words: Vowel length lengthen and Vowel Length Shorten, Consonant confusion and Vowel confusion, Syllabification, Addition and Omission, and Others. Unanswered words were also counted as errors. All categories violated normality with a *p*-value less than .001.

Real Word: Vowel Length Lengthen and Vowel Length Shorten

Kruskal-Wallis

	χ^2	df	p
RW.VLL.percent	8.33	2	0.016
RW.VLS.percent	9.77	2	0.008

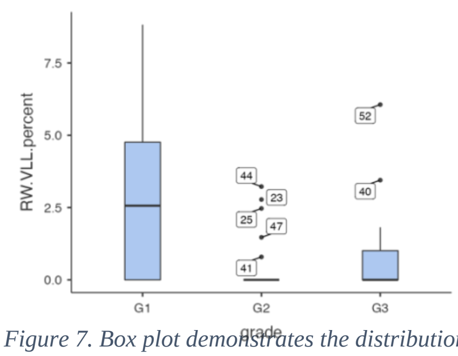

Figure 7. Box plot demonstrates the distribution of percentages in VLL Grade 1, Grade2 and Grade 3.

Pairwise comparisons - RW.VLL.percent

		W	p
G1	G2	-3.74	0.022
G1	G3	-3.10	0.072
G2	G3	2.02	0.326

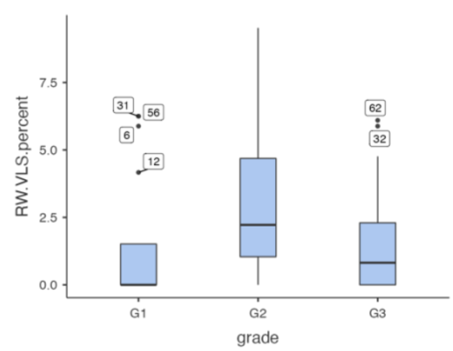

Figure 8. Box plot demonstrates the distribution of percentages in VLS across Grade 1, Grade2 and Grade 3.

Pairwise comparisons - RW.VLS.percent

		W	p
G1	G2	3.41	0.042
G1	G3	1.34	0.609

Pairwise comparisons - RW.VLS.percent

		W	p
G2	G3	-3.93	0.015

Assumption of normality was violated for both cumulative percentages of Vowel Length Lengthen errors (*W* = 0.79, *p*<.001) and Vowel Length Shorten errors (*W* = 0.86, *p*<.001) in reading real words, thus assuming that the data were not derived from a population with the same distribution. We continued with the Kruskal-Wallis test which examines the data ranks across the three grades. The table shows that there are statistically significant differences in Vowel Length Lengthen errors across all three grades ($\chi2(2) = 8.33$, *p* = .016). The post-hoc test DSCF pairwise comparison reports that, for Vowel Length Lengthen errors, there is a statistically significant difference of errors between Grade 1 and Grade 2 (*W* = -3.74, *p* = .022), but not between Grade 1 and Grade 3 (*W* = -3.10, *p* = .072) and Grade 2 and Grade 3 (*W* = 2.02, *p* = .326). The box plot on Figure 7 shows variety in Grade 1 compared to the other two grades, especially Grade 2 which have many outliers instead, showing the significant differences between the two grades.

The means of percentages of Vowel Length Lengthen errors in Grade 1 is 2.63, 0.43 in Grade 2, and 0.63 in Grade 3. Kruskal-Wallis reported that there are also statistically significant differences in VLS errors across all grades ($\chi2(2) = 9.77$, *p* = .008). DSCF test shows that there are also significant differences between Grade 1 and Grade 2 (*W* = 3.41, *p* = .042) and Grade 2 and Grade 3 (*W* = -3.93, *p* = .015), but not between Grade 1 and Grade 3 (*W* = 3.41, *p* = .609). The mean percentage of VLS errors is 1.42 in Grade 1, 2.97 in Grade 2 and 1.35 in Grade 3. Grade 2 seems to have more dispersion in data, showing that there are differences in error scores between Grade 1 and Grade 2 as well as Grade 2 and Grade 3.

Real Word: Consonant Substitution and Vowel Substitution

Kruskal-Wallis

	χ^2	df	p
pse.CS	4.79	2	0.091
pse.VS	11.76	2	0.003

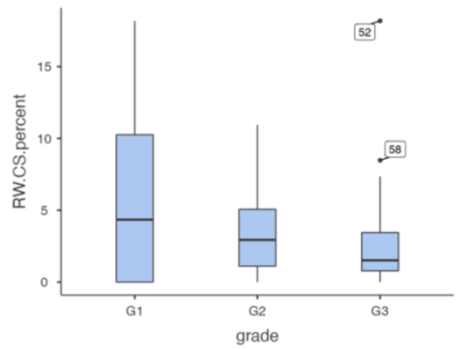

Figure 9. Box plot demonstrates the distribution of percentages in CS across Grade 1, Grade2 and Grade 3.

Pairwise comparisons - RW.CS.percent

		W	p
G1	G2	-1.10	0.718
G1	G3	-2.06	0.311
G2	G3	-2.54	0.170

Pairwise comparisons - RW.VS.percent

		W	p
G1	G2	-3.85	0.018
G1	G3	-4.91	0.001
G2	G3	-1.03	0.746

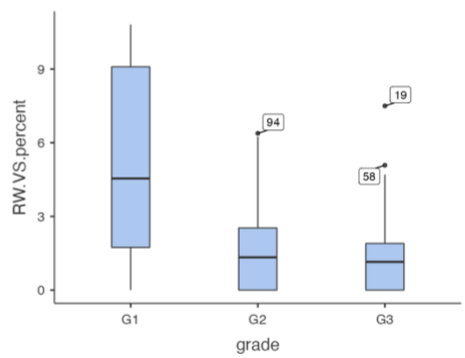

Figure 10. Box plot demonstrates the distribution of percentages in VS across Grade 1, Grade2 and Grade 3.

Kruskal-Wallis test reports that there is not a statistically significant difference in Consonant Substitution errors across all grades ($\chi2(2) = 4.32$, $p = .115$). DSCF also reports no statistically significant differences across all grades: Grade 1 ($M = 5.93$) and Grade 2 ($M = 3.53$) ($W = -1.10$, $p = .718$), Grade 1 and Grade 3 ($M = 2.46$) ($W = -2.06$, $p = .311$), and Grade 2 and Grade 3 ($W = -2.54$, $p = .170$). The box plot shows a bigger distribution in Grade 1 compared to the other grades, and Grade 3 has an outlier showing that a participant has a way higher error score than others.

On the other hand, the Kruskal-Wallis test showed that there was a significant difference between Vowel Substituion errors across all grades $\chi2(2) = 12.72$, $p = .002$), and the DSCF test shows that there was a significant difference in errors between Grade 1 (*M* = 5.18) and Grade 2 (*M* =1.88) (*W* = -3.85, *p* = .018) and Grade 1 and Grade 3 (*M* = 1.45) (*W* = -4.91, *p* = .001), but not found between Grade 2 and Grade 3 (*W* = -1.03, *p* = .746). The data were more distributed in Grade 1 compared to the other two grades which supports the results of significant differences between Grade 1 and Grade 2 as well as Grade 1 and Grade 3.

Real Word: Syllabification

Kruskal-Wallis

	χ^2	df	p
RW.SYB.percent	4.65	2	0.098

Pairwise comparisons - RW.SYB.percent

		W	p
G1	G2	-2.95	0.092
G1	G3	-2.04	0.319
G2	G3	1.71	0.450

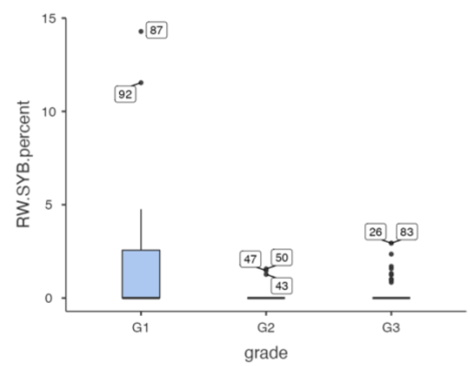

Figure 11. Box plot demonstrates the distribution of percentages in SYB across Grade 1, Grade2 and Grade 3.

Kruskal-Wallist test reports that there are no significant differences across all grades ($\chi2(2) = 4.65$, *p* = .098). The DSCF post-hoc test also shows no significant differences between Grade 1 (*M* = 2.31) and Grade 2 (*M* = 0.17) (*W* = -2.95, *p* = .092), Grade 1 and Grade 3 (*M* = 0.38) (*W* = -2.04, *p* = .319), and Grade 2 and Grade 3 (*W* = 1.71, *p* = .450). Figure 11 shows outliers in Grade 1 and scattered scores across all grades.

Real word: Addition and Omission

Kruskal-Wallis

	χ^2	df	p
RW.ADD.percent	4.26	2	0.119
RW.OMI.percent	8.86	2	0.012

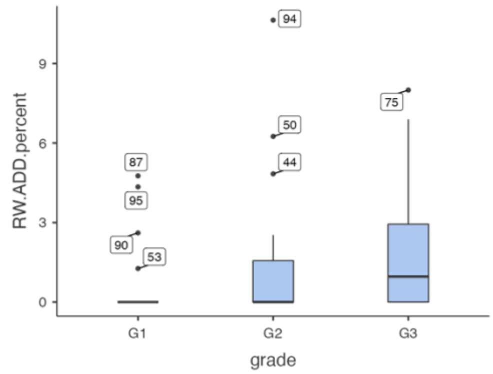

Figure 12. Box plot demonstrates the distribution of percentages in ADD across Grade 1, Grade2 and Grade 3.

Pairwise comparisons - RW.ADD.percent

		W	p
G1	G2	1.73	0.438
G1	G3	2.87	0.106
G2	G3	1.19	0.676

Pairwise comparisons - RW.OMI.percent

		W	p
G1	G2	2.77	0.123
G1	G3	4.21	0.008
G2	G3	1.44	0.568

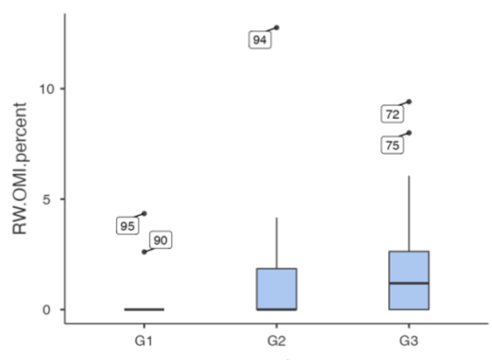

Figure 13. Box plot demonstrates the distribution of percentages in OMI across Grade 1, Grade2 and Grade 3.

There were no significant differences in errors for Addition between all grades ($\chi 2(2) = 4.26$, *p* = .119), also shown by the DSCF test and means: Grade 1 (*M* = 0.76) and Grade 2 (*M* = 1.44) (*W* = 1.73, *p* = .438), Grade 1 and Grade 3 (*M* = 1.66) (*W* = 2.87, *p* = .106), and Grade 2 and Grade 3 (*W* = 1.19, *p* = .676). However, there were differences in errors for Omission ($\chi 2(2) = 8.86$, *p* = .012), but only for the comparison between Grade 1 (*M* = 0.41) and Grade 3 (*M* = 1.74) were significant (*W* = 4.21, *p* = .008), while the comparison between Grade 1 and Grade 2 (*M* =1.52)

(*W* = 2.77, *p* = .123) and Grade 2 and Grade 3 (*W* = 1.44, *p* = .568) was not. Box plots show that scores are scattered in Grade 1 with many outliers across all grades. The box plot shows little data in Grade 1, also found in the mean score which is 0.41 compared to the mean score in Grade 3 which is 1.44, showing that there is a significant difference between Grade 1 and Grade 3 in Omission errors.

Real Word: Other errors

Kruskal-Wallis

	χ^2	df	p
RW.OTH.percent	2.51	2	0.285

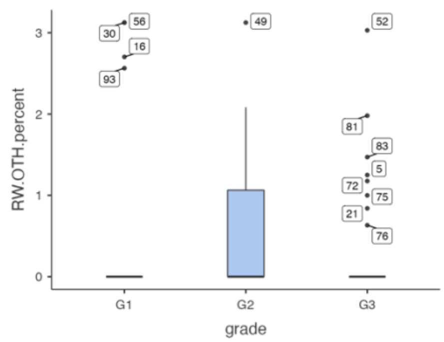

Figure 14. Box plot demonstrates the distribution of percentages in OTH across Grade 1, Grade2 and Grade 3.

Pairwise comparisons - NW.OTH.percent

		W	p
G1	G2	-0.35	0.967
G1	G3	-2.32	0.229
G2	G3	-2.76	0.125

The category of other error mistakes violated the assumption of normality (*W* = 0.69, *p*<.001). Kruskal-Wallis test reports that there are no significant differences across grades (χ2(2) = 4.87, *p* = .088), supported by the DSCF test and the grades' means of percentages compared to each other: Grade 1 (*M* =0.68) and Grade 2 (*W* = -0.35, *p* = .967), Grade 1 and Grade 2 (*M* = 0.49) (*W* = -2.32, *p* = .229), and Grade 2 and Grade 3 (*M* = 0.21*)* (*W* = -2.76, *p* = .125).

6.3.2 Nonword errors

Descriptives

Descriptives

			Shapiro-Wilk	
	Mean	SD	W	p
NW.nErr	18.25	12.20	0.80	< .001
NW.VLL.percent	2.81	2.94	0.85	< .001
NW.VLS.percent	4.58	4.39	0.89	< .001
NW.CS.percent	5.89	5.26	0.89	< .001
NW.VS.percent	4.99	3.61	0.93	< .001
NW.SYB.percent	3.66	4.28	0.77	< .001
NW.ADD.percent	1.63	2.58	0.69	< .001
NW.OMI.percent	2.01	2.58	0.78	< .001
NW.LEX.percent	0.32	0.81	0.46	< .001
NW.OTH.percent	1.29	2.48	0.58	< .001
NW.UA.percent	4.14	11.29	0.40	< .001

Reading errors from all grades were also found in all categories, including Lexicalisation, when reading the Nonword lists. All categories violated normality with a *p*-value less than .001.

Nonword: Vowel Length Lengthen and Vowel Length Shorten

Kruskal-Wallis

	χ^2	df	p
NW.VLL.percent	6.43	2	0.040
NW.VLS.percent	15.00	2	< .001

Pairwise comparisons - NW.VLL.percent

		W	p
G1	G2	1.72	0.446
G1	G3	-0.49	0.937
G2	G3	-3.72	0.023

Figure 15. Box plot demonstrates the distribution of percentages in VLL (from Nonword errors) across Grade 1, Grade2 and Grade 3.

Pairwise comparisons - NW.VLS.percent

		W	p
G1	G2	4.80	0.002
G1	G3	3.74	0.022
G2	G3	-3.47	0.038

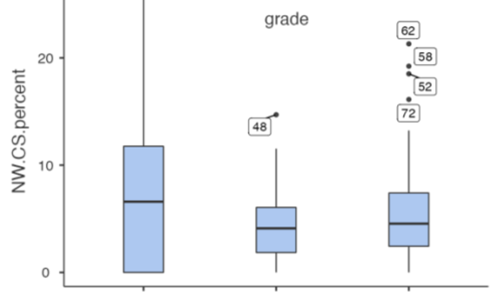

Figure 16. Box plot demonstrates the distribution of percentages in VLS (from Nonword errors) across Grade 1, Grade 2 and Grade 3.

The non-parametric ANOVA test Kruskal-Wallis was therefore conducted. The test reports there are statistically significant differences of Vowel Length Lengthen errors across the three grades ($\chi 2(2) = 6.43$, $p = .040$). However, the DSCF pairwise comparisons show that there are no significant differences between Grade 1 ($M = 2.75$) and Grade 2 ($M = 4.10$) ($W = 1.72$, $p = .446$) and Grade 1 and Grade 3 ($M = 2.22$) ($W = -0.49$, $p = .937$), but found difference in scores between Grade 2 and Grade 3 ($W = -3.72$, $p = .023$). Data in Grade 2 seems to be dispersed with many outliers, but within the range of Grade 1 compared to Grade 3 which was reported to be significant.

On the other hand, errors in Vowel Length Shorten showed significant differences across grades ($\chi 2(2) = 15.00$, $p<.001$). The DSCF test supports this by showing significant differences between Grade 1 ($M = 2.11$) and Grande 2 ($M = 7.33$) ($W = 4.80$, $p = .002$), Grade 1 and Grade 3 ($M = 4.08$) ($W = 3.74$, $p = .022$), and Grade 2 and Grade 3 ($W = -3.47$, $p = .038$). The box plot shows a bigger range of errors in Grade 2 compared to Grade 1 and 3 which have outliers. This can suggest differences between grades as reported by DSCF.

Nonword: Consonant Substitution and Vowel Substitution

Kruskal-Wallis

	χ^2	df	p
NW.CS.percent	1.07	2	0.586
NW.VS.percent	0.58	2	0.747

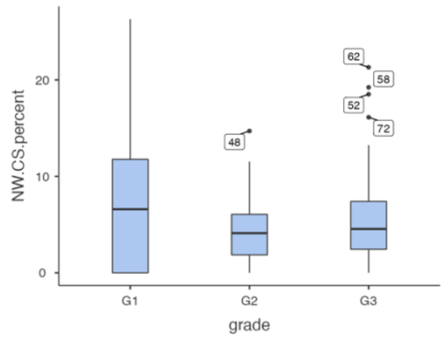

Figure 17. Box plot demonstrates the distribution of percentages in CS (from Nonword errors) across Grade 1, Grade2 and Grade 3.

Pairwise comparisons - NW.CS.percent

		W	p
G1	G2	-1.42	0.573
G1	G3	-0.81	0.833
G2	G3	0.95	0.778

Pairwise comparisons - NW.VS.percent

		W	p
G1	G2	0.85	0.818
G1	G3	1.06	0.735
G2	G3	0.14	0.995

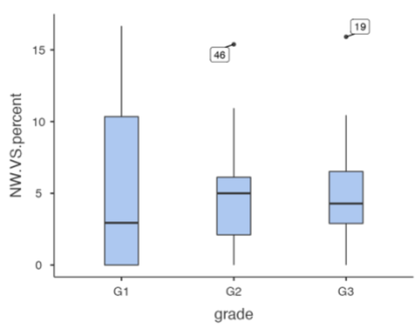

Figure 17. Box plot demonstrates the distribution of percentages in VS (from Nonword errors) across Grade 1, Grade2 and Grade 3.

Kruskal-Wallis reports that there are no significant differences in Consonant Substitution across all grades: ($\chi 2(2) = 1.07$ $p = .586$), which is also shown by the DSCF test: Grade 1 (*M* = 7.59) and Grade 2 (*M* = 4.84) (*W* = -1.42, *p* = .573), Grade 1 and Grade 3 (*M* = 5.84) (*W* = -0.81, *p* = .833), and Grade 2 and Grade 3 (*W* = 0.95, *p* = .778) did not have significant differences. Furthermore, Vowel Substitution also did not have statistically significant differences across the three grades ($\chi 2(2) = 0.58$, $p = .747$), also shown by the comparisons between Grade 1 (*M* = 4.99) and Grade 2 (*M* = 4.99) (*W* = 0.85, *p* = .818), Grade 1 and Grade 3 (*M* = 4.98) (*W* = 1.06, *p* = .735), and Grade 2 and Grade 3 (*W* = 0.14, *p* = .995). As shown on both box plots, Grade 1 had a wider range of nonword errors compared to the two other grades, however, Grade 3 had many outliers in Consonant Substitution errors.

Nonword: Syllabification

Kruskal-Wallis

	χ^2	df	p
NW.SYB.percent	0.24	2	0.885

Pairwise comparisons - NW.SYB.percent

		W	p
G1	G2	-0.48	0.939
G1	G3	-0.15	0.994
G2	G3	0.68	0.882

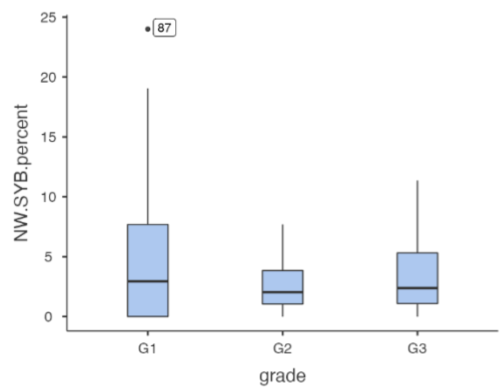

Figure 18. Box plot demonstrates the distribution of percentages in SYB (from Nonword errors) across Grade 1, Grade2 and Grade 3.

Kruskal-Wallis test reported ($\chi 2(2) = 0.24$, $p = .885$), meaning there are no significant differences in errors across grades: Grade 1 (*M* = 5.71) and Grade 2 (*M* =2.71) (*W* = -0.48, *p* = .939), Grade 1 and Grade 3 (*M* = 3.44) (*W* = -0.15, *p* = .994), and Grade 2 and Grade 3 (*W* =

0.68, p = .882). The box plot shows a slightly wider range of errors for Grade 1 along with an outlier compared to Grade 2 and Grade 3.

Nonword: Addition and Omission

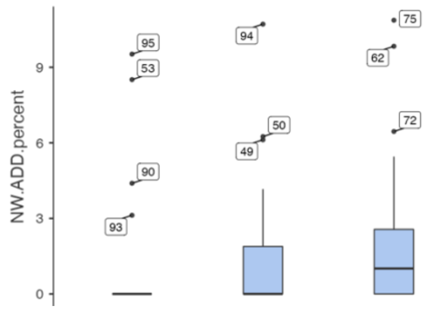

Figure 19. Box plot demonstrates the distribution of percentages in ADD (from Nonword errors) across Grade 1, Grade2 and Grade 3.

Kruskal-Wallis

	χ^2	df	p
NW.ADD.percent	2.26	2	0.324
NW.OMI.percent	8.68	2	0.013

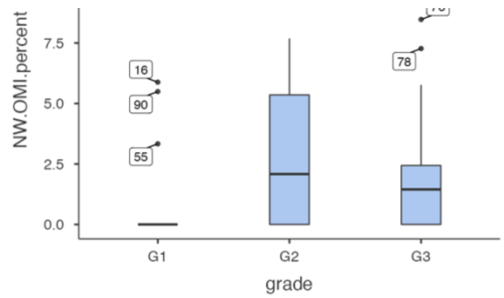

Figure 20. Box plot demonstrates the distribution of percentages in OMI (from Nonword errors) across Grade 1, Grade2 and Grade 3.

Pairwise comparisons - NW.ADD.percent

		W	p
G1	G2	1.04	0.744
G1	G3	2.01	0.331
G2	G3	1.12	0.710

Pairwise comparisons - NW.OMI.percent

		W	p
G1	G2	3.52	0.035
G1	G3	3.84	0.018
G2	G3	-1.54	0.522

Kruskal-Wallis test reports that there are no significant differences in Addition errors across the grades ($\chi2(2) = 2.26$, $p = .324$) shown by the DSCF test: Grade 1 ($M = 1.50$) and Grade 2 ($M = 1.56$) ($W = 1.04$, $p = .744$), Grade 1 and Grade 3 ($M = 1.71$) ($W = 2.01$, $p = .331$), and Grade 2 and Grade 3 ($W = 1.12$, $p = .710$). Errors in Omission, however, seemed to have statistically significant differences across grades ($\chi2(2) = 8.68$, $p = .013$). The pairwise comparison shows significant differences between Grade 1 ($M = 0.87$) and Grade 2 ($M = 2.73$) ($W = 3.52$, $p = .035$) and Grade 1 and Grade 3 ($M = 2.03$) ($W = 3.84$, $p = .018$), but not between Grade 2 and Grade 3 ($W = -1.54$, $p = .522$). The boxplot for Addition shows many outliers in all grades, suggesting higher error percentages in all three. Due to this, there are no big differences in percentages between grades. Grade 1 and Grade 3 both had many outliers in errors of Omission, showing that Grade 3 had higher percentages of reading nonwords compared to Grade 1. It was not significant between Grade 2 and Grade 3 possibly because of a wider range of errors in Grade 2 without outliers.

Nonword: Lexicalisation

Kruskal-Wallis

	χ^2	df	p
NW.LEX.percent	15.13	2	< .001

Pairwise comparisons - NW.LEX.percent

		W	p
G1	G2	4.12	0.010

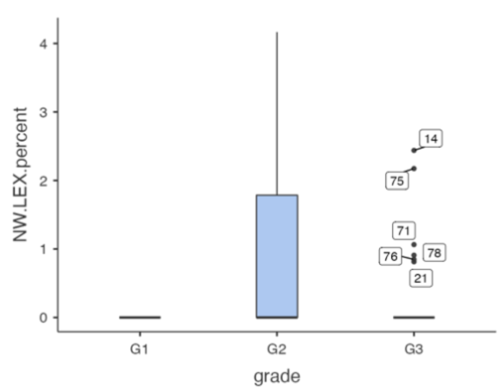

Figure 21. Box plot demonstrates the distribution of percentages in LEX (from Nonword errors) across Grade 1, Grade2 and Grade 3.

Pairwise comparisons - NW.LEX.percent

		W	p
G1	G3	2.03	0.321
G2	G3	-4.39	0.005

Lexicalisation refers to reading a nonword as a real word with phonological similarities and is therefore included in the analyses of nonword errors. Lexicalisation violated the assumption of normality (*W* = 0.71, *p*<.001). Kruskal-Wallis reported that there are statistically significant differences in Lexicalisation errors across the three grades ($\chi 2(2) = 15.13$, *p*<.001). DSCF results show that there were significant differences between Grade 1 (*M* = 0.00) and Grade 2 (*M* = 1.27) (*W* = 4.12, *p* = .010) and Grade 2 and Grade 3 (*M* = 0.49) (*W* = -4.39, *p* = .005), but not between Grade 1 and Grade 3 (*W* = 2.03, *p* = .321). The box plot shows that the error mistakes across grades were significantly different with no errors in Grade 1, significantly higher error scores in Grade 2 and dispersed error scores in Grade 3.

Nonword: Other errors

Kruskal-Wallis

	χ^2	df	p
NW.OTH.percent	4.87	2	0.088

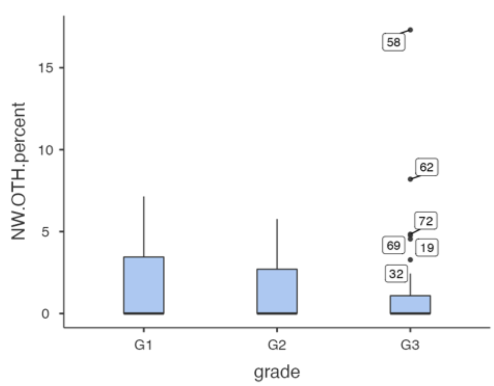

Figure 22. Box plot demonstrates the distribution of percentages in OTH (from Nonword errors) across Grade 1, Grade2 and Grade 3.

Pairwise comparisons - NW.OTH.percent

		W	p
G1	G2	-0.35	0.967

Pairwise comparisons - NW.OTH.percent

		W	p
G1	G3	-2.32	0.229
G2	G3	-2.76	0.125

Other errors include uncategorised reading mistakes. The Kruskal-Wallis table shows that there were no significant differences in uncategorised errors across grades ($\chi 2(2) = 4.87$, $p = .088$). Comparing pairs of all grades also shows that there were no significant differences between Grade 1 and Grade 2 ($W = -0.35$, $p = .967$), Grade 1 and Grade 3 ($W = -2.32$, $p = .229$), and Grade 2 and Grade 3 ($W = -2.76$, $p = .125$). The box plot does not show any big differences in error scores but shows that in Grade 3 the error scores were more spread out across participants.

7 Discussion

This study aimed to identify and categorise the reading mistakes of Norwegian children in primary grades. The categories identified in this present study are Vowel Length Confusion (VCL, both shortened, VLS, and lengthened, VLL), Consonant Substitution (CS) and Vowel Substitution (VS), Syllabification (SYB), Addition (ADD) Omission (OMI), Lexicalisation (LEX), and Other (OTH) reading mistakes. Unanswered were marked as incorrectly, but it is not included as one of the main categories. Some of these categories are similar to the four categories found in Nergård-Nilsen's (2005) study that partially aimed to examine the components of reading mistakes based on Norwegian orthography by examining four children with dyslexia. Their four categories are vowel length confusions, vowel confusions, consonant cluster misreading, and other errors.

The first category, Vowel Length Confusion, consists of two sub-categories, Vowel Length Lengthen and Vowel Length Shorten. Awareness of vowel length rules in Norwegian can demonstrate the reader's level of reading development. In Norwegian orthography, the vowels, *a, e, i, o, u, y, æ, ø, å,* are greatly affected by the consonants surrounding them. Consonant clusters can determine the length of the vowel; if the preceding consonant is a single letter, the vowel sound is lengthened, but if the following are consonant clusters, the vowel sound shortens (Hillestad, et al., n.d.). For instance, the vowel *ø in the word søt* (/søːt/, cute) is dragged out compared to the *ø* in the word *søtt* (/sœʰtː/, sweet) which has a shorter vowel sound. In this study, a common mistake was pronouncing the word *vil* (/ʋɪl/, want) with a long vowel. Although the following is a single consonant, the vowel sound is short and does not follow the rule of long vowel sound, with the same pronunciation as the word *vill* ((/ʋɪl/, wild) which ends with a double consonant. For nonwords, *døke*(/dø:ke/) was commonly mistaken as *døkke*, pronounced with a shorter vowel. Similar findings were found in four children with dyslexia who mispronounced vowels which led to confusion of unrelated words (Nergård-Nilssen, 2005).

Consonant Substitution and Vowel Substitution refer to when the reader would confuse the consonant or vowel for another. Usually, they would substitute it for a letter that is visually or phonologically similar. The word *bil* (car) was often read as *dil* where the child switched the letter *b* to *d*, which are visually similar. Similarly, vowels were also switched like in the nonword *go* which was pronounced as *gå*. In this case, it might be due to phonological similarities, or the child simply read the nonword as the real word *gå* (walk). Nergård-Nilssen (2005) found that the most common mistakes were between the vowels o and å, e and æ, and o and u, suggesting that there are irregularities in the Norwegian vowels.

Addition and Omission refer to removing or adding letters or phonemes in a word. These have been categorised as errors in previous studies as well (Weber, 1970; Leu Jr., 1982; Marinelli et al., 2023). Mispronounced words were put under these categories if the reader removed a letter in a word, inserted a letter, moved the placement of letters, or substituted a letter in the word (Marinelli et al., 2023). This study mostly focused on whether the reader removed or added a letter or sound. Common mistakes were reading the word *stort* (big, referring to a noun) to *stor* (big), and adding *n* in the word *noe* (something) which turned into *noen* (someone). It can be speculated that these mistakes were common because the incorrectly read word is a real word that is relevant to the target word. For nonwords, readers often removed the *l* in the nonword *færl* or added an *i*. The results would either be *fær* or *færli*, either because the child struggled to combine a consonant cluster (or roll their tongue), or they added an *i* to make it sound like the word *farlig* (/fa:rli/, dangerous).

Syllabification refers to when the reader tries to decode the word by dividing the word into syllables, such as the word *fugl* (/fu:l/, bird) would be syllabised as *fug-gel*. Not only did the reader add a syllable, but they also pronounced the letter *g* which is silent in the actual word. Syllabification often occurs when the word consists of consonant clusters and the word *fugl* has the structure of CVCC. The most common mistake found in all three grades regarding nonwords was the mispronunciation of *skjad* (/ʃa:d/), which has a CCCVC structure. This nonword was read in different ways such as *sk-yad, s-kadd, s-kk-yad, sa-ka-yad,* or *s-k-yab*. The readers chose to decode the word by dividing it into two or more syllables, which can seem like the child is using the sounding-out behaviour to progressively get to the target word (Marinelli et al., 2023). Different methods were used to syllabise the word, those who broke the word into *sk-yad* separated the consonant cluster away from the rest of the word, while those who read it as *ska-yad* also separated but added a vowel sound in the first syllable. Some readers chose to decode the word phoneme by phoneme and separated the consonants within the clusters, such as *s-kk-yad*. Similar findings were seen in Nergård-Nilssen's (2005) study where the word *sjø* (sea) was broken up into individual sounds (/s/, /j/, /3/), which suggests that the readers relied on their grapheme-phoneme awareness to decode a complex word. This error totalled 21.4 to 40,9% of all reading mistakes, suggesting that the children were unable to use the orthographic rules when they had to read words with no contextual clues (Nergård-Nilssen, 2005). Lexicalisation was included as a category solely for nonword reading. It refers to how the target word, usually a nonword, ends up being read as a real word. This can happen due to similarities in letters or spoken sounds (Marinelli et al., 2003). In this present study, the nonword *kjerk* (/Ç:ærk/) was read as *kjekk* (/çɛk/, handsome). The nonword turned into a real word by swapping the letter *r* to a *k*, which on print looks similar. However, their pronunciations are not that similar, meaning the child was unable to decode the letter-sound relationship when it came to the vowel *e* as it was pronounced as /æ:/ instead of /e:/ when combined with the letter *r*, as seen in the Norwegian word *er* (is) which is pronounced as /æ:r/. Another nonword is *kæse* (/kæ:se/) which was read as the Norwegian word *kjæreste* (/kjæ:reste/, beloved). In this case, the reader added syllables to the nonwords to form it into a real word. Both words have similar letters, but compared to the previous example, these words have similar pronunciations because of the /æ:/ sound.

The last category, others, consists of reading mistakes that were not categorised due to a few numbers of those mistakes. Many of the reading mistakes are restricted to one category only as most of them could fit in more than one category. Some of the examples above, such as *kæse* to *kjæreste* could be categorised in Lexicalisation and Addition, or the mispronunciations of *fugl* could be included in Syllabification, Addition or Consonant Substitution as the reader was not aware of the silent consonant and added a hard *g* sound instead. Other types of mistakes were words that were read phonologically correctly by rules, but were not correct pronunciations of real words. For instance, the word *verk* (/væ:rk/, work or piece) was sometimes read as /ve:rk/, as the letter *e* is pronounced as /e:/ in the alphabet. The word *kyss* (/kjys/, kiss) was also read with a hard *k* a few times as the letter is pronounced as /ko:/ and not /kj/. This can be confusing for readers who are not yet familiar with the rules between consonants and vowels, as the pronunciation of the letter *k* is hard when the following vowels are a, o, u, å, but becomes soft when the following vowels are, for example, e, i, and y.

In general, the results suggest that Grade 3 scores higher than Grade 1 and Grade 2 in reading both real words and nonwords.
The ANOVA test for lexicalisation shows that Grade 1 had no lexicalisation mistakes compared to the two other grades. This is because children at this age either did not get far enough to read more complex words like Grade 2 and Grade 3, or they do not have enough grapheme-phoneme awareness to make associations between nonwords and real words, thus they do not have other words to confuse the nonwords with.

One big limitation of this study is the different number of participants in each grade. Grade 1 had the least participants and also the least number of words read as children in Grade 2 and Grade 3 have developed reading skills further than them. This made it hard to compare their scores as some children in Grade 3 read many more words than the average child in Grade 1. There is therefore a big difference in the portion of mistakes due to the different number of participants.

For future research, it would be insightful to extend the sample to higher grades to examine the reading mistakes of older children.

8 References

Adsett, C. R., & Marchand, Y. (2010). Syllabic complexity: A computational evaluation of nine European languages. *Journal of Quantitative Linguistics*, *17*(4), 269–290. https://doi.org/10.1080/09296174.2010.512161

Başar, M., & Batur, Z. (2018). Examination of the students' mistakes in oral reading. *International Electronic Journal of Eleme*

Byrne, B. (2014). The foundation of literacy: The child's acquisition of the alphabetic principle. Psychology Press.

Byrne, B., & Fielding-Barnsley, R. (1989). Phonemic awareness and letter knowledge in the child's acquisition of the Alphabetic Principle. *Journal of Educational Psychology*, *81*(3), 313–321. doi:10.1037/0022-0663.81.3.313

Campbell, D. T., & Cook, T. D. (1979). Quasi-experimentation. *Chicago, IL: Rand Mc-Nally*.

Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education*, *4*(1). https://doi.org/10.1080/13603110600574322

Das, J. P. (2009). Reading difficulties and dyslexia: An interpretation for teachers. Sage Publications.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4

Ehri, L. C., & McCormick, S. (1998). Phases of word learning: Implications for instruction with delayed and disabled readers. *Reading & Writing Quarterly*, *14*(2), 135163. https://doi.org/10.1080/1057356980140202

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, *36*(3), 250–287. https://doi.org/10.1598/rrq.36.3.2

Erdem, A. (2017). Examination of Oral Reading Mistakes of Primary School Students. *Educational Research and Practice, 179.*

Ferguson, L. (2004). External validity, generalizability, and knowledge utilization. *Journal of Nursing Scholarship*, *36*(1), 16-22.

Findley, M. G., Kikuta, K., & Denly, M. (2021). External validity. *Annual Review of Political Science*, *24*, 365-393.

Frith, U. (1986). A developmental framework for developmental dyslexia. *Annals of Dyslexia*, *36*(1), 67–81. https://doi.org/10.1007/bf02648022

Goswami, U. (1999). The relationship between phonological awareness and orthographic representation in different orthographies. Learning to read and write: A cross-linguistic perspective, 134-156.

Goswami, U., Ziegler, J. C., Dalton, L., & Schneider, W. (2003). Nonword reading across orthographies: How flexible is the choice of reading units? *Applied Psycholinguistics*, *24*(2), 235–247. doi:10.1017/s0142716403000134

Hagtvet, B. E., Helland, T., & Lyster, S. A. H. (2006). Literacy acquisition in Norwegian. *Handbook of orthography and literacy*, 15-30.

Harrison, C. (2004). Understanding reading development. Sage.

Hendriks, A. W., & H.J. Kolk, H. H. J. (1997). Strategic control in developmental dyslexia. *Cognitive Neuropsychology*, *14*(3), 321–366. doi:10.1080/026432997381510

Hillestad, B., Husby, O., Øvregaard, Å., & Robbins, S. (n.d.). *Norwegian on the web - NTNU*. Pronunciation. http://www.hf.ntnu.no/now/hardcopies/Pronunciation.pdf

Labov, W., & Baker, B. (2010). What is a reading error?. *Applied Psycholinguistics*, *31*(4), 735-757.

Larsen, L., Hjetland, H. N., & Schauber, S. K. (2022). A longitudinal investigation of letter-name knowledge in a semi-transparent orthography. Scandinavian Journal of Educational Research, 1-15

Lervåg, A., Bråten, I., & Hulme, C. (2009). The cognitive and linguistic foundations of early reading development: A Norwegian latent variable longitudinal study. *Developmental Psychology*, *45*(3), 764–781. https://doi.org/10.1037/a0014132

Leu Jr, D. J. (1982). Oral reading error analysis: A critical review of research and application. Reading Research Quarterly, 420-437.

Marinelli, C. V., Romani, C., McGowan, V. A., Giustizieri, S., & Zoccolotti, P. (2023). Characterization of reading errors in languages with different orthographic regularity: An Italian–English comparison. *Journal of Cultural Cognitive Science*, *7*(2), 95–120. doi:10.1007/s41809-023-00119-1

Murray, L. (2018). *Oral vocabulary and reading new words: examining the mispronunciation correction process* (Doctoral dissertation, Macquarie University).

Nation, K., & Cocksey, J. (2009). The relationship between knowing a word and reading it aloud in children's word reading development. *Journal of Experimental Child Psychology*, *103*(3), 296-308.

Navarro, D., & Foxcroft, D. (2019). Learning statistics with jamovi: A tutorial for psychology students and other beginners (Version 0.70).

Norwegian orthographic analyzer. Institutt for spesialpedagogikk. (2023, June 1). https://www.uv.uio.no/isp/om/oslo-spesialpedagogikk-og-laeringslab/utlan-og-tekniske-tjenester/norwegian-orthographic-analyzer/

Nunes, T., Bryant, P., & Bindman, M. (1997). Morphological spelling strategies: developmental stages and processes. *Developmental psychology*, *33*(4), 637.

Pagliuca, G., Monaghan, P., & McIntosh, R. (2008). Variable vulnerability of words to visual impairment: exploring grain-size effects in reading. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 30, No. 30).

Perea, M., & Estévez, A. (2008). Transposed-letter similarity effects in naming pseudowords: Evidence from children and adults. *European Journal of Cognitive Psychology*, *20*(1), 33-46.

References

Seidenberg, M. S. (2007). *Connectionist models of reading* (pp. 235-50). Oxford Handbook of Psycholinguistics.

Seymour, P. H. (2005). Early reading development in European Orthographies.

Seymour, P. H., Aro, M., Erskine, J. M., & Collaboration with COST Action A8 Network. (2003). Foundation literacy acquisition in European orthographies. *British Journal of psychology*, *94*(2), 143-174.

Simoncelli, E. P., Paninski, L., Pillow, J., & Schwartz, O. (2004). The cognitive neurosciences. Chapter, 23, 327-338.

The jamovi project (2022). *jamovi*. (Version 2.3) [Computer Software]. Retrieved from https://www.jamovi.org.

Torgeson, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test Review: Test of Word Reading Efficiency (TOWRE)*. Pro-ed.

Universitetet i Oslo . (2023, June 1). *Norwegian Orthographic Analyzer*. Institutt for spesialpedagogikk . https://www.uv.uio.no/isp/om/oslo-spesialpedagogikk-og-laeringslab/utlan-og-tekniske-tjenester/norwegian-orthographic-analyzer/

Wagner, R. K. (2008). Learning to read: The importance of assessing phonological decoding skills and sight word knowledge. Scholastics Phonics Inventory.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. Psychological bulletin, 101(2), 192.

Weber, R.-M. (1970). A linguistic analysis of first-grade reading errors. *Reading Research Quarterly*, *5*(3), 427. https://doi.org/10.2307/747079

Ziegler, J. C., Perry, C., & Coltheart, M. (2003). Speed of lexical and nonlexical processing in French: The case of the regularity effect. *Psychonomic Bulletin & Review*, *10*(4), 947–953. doi:10.3758/bf03196556

9 Appendices

Non-ord	Reelle ord, eksempel	Østnorsk uttale/regler
Pe	Se	Pe: med lang e
Iv	Av	I:v med lang i
Go	Ro	Go: med lang o
Bå	Så	Bå: med lang å
Um	Om	Um: med trykk på m
Ky	Kyss eller kypros	Ky: med lang y Bokstavkombinasjonen <ky> blir uttalt çʏ, som i kyss, kyst, kypros.
Fæ	Få	Fæ: med lang æ
Noff	Voff eller loff	Nof: kan uttales som voff med å eller o som i loff
Tum	Dum eller sum	To:m med trykk på m Bokstaven <u> uttales ʊ, som i ordene dum, skuff. Her kan man også uttale ʉ, som i sum, slum.
Pitt	Litt	Pit: med dobbel t
Mof	Kos	Mo:f med lang o
Rog	Bog	Ro:g med lang o
Kib	Siv	Ki:b med lang i
Fav	Hav	Fa:v med lang a
Skjad	Skjev	ʃa:d med lang a Bokstavkombinasjonen <skj> blir uttalt ʃ, som i skjev
Kæse	Pøse	Bæ:se med lang æ
Døke	Søke	Dø:ke med lang ø
Lieb	Liam	Li:eb med lang i, som i navnet Liam
Færl	Perle	Fæ:l med lang æ. Bokstavkombinasjonen <rl> blir uttalt l, som i perle
Turp	Slurp	Tu:rp med lang u
Kjerk	Kjær	Ç:ærk med trykk på ç Bokstavkombinasjonen <kj> blir uttalt ç, som i kjær. Bokstaven <e> foran <r> uttales æ, som i verden, verpe
Jåst	Låst	Lå:st med lang æ
Stru	Strø	Stru: med lang u
Veif	Veiv	Væ͜ɪ:f med trykk på æ͜ɪ Bokstavkombinasjonen <ei> blir uttalt æ͜ɪ, som i ordet stein, veive. Det godtas også med

		uttalen «ei» da noen dialekter uttaler bokstavkombinasjonen slik, som for eksempel personer fra Trøndelag.
Natsj	Dusj	Batʃ: med trykk på ʃ Bokstavkombinasjonen <sj> blir uttalt ʃ, som i ordet dusj.
Grakk	Krakk	Grak: med dobbel k
Frup	Glup	Fru:p med lang u
Rink	Ring	Ri:ŋ:k, med lang i. Bokstavkombinasjonen <nk> blir uttalt ŋ:k, som i bank.
Losp	Kost	Los:p med trykk på s
Mart	Bart	Baʈ: med trykk på ʈ Bokstavkombinasjonen <rt> blir uttalt ʈ, som i ordet bart
Stark	Sterk	Star:k med trykk på r
Gjef	Gjøk	Je:f med lang e Bokstavkombinasjonen <gj> blir uttalt j, som i gjerde
Kotter	Votter	Kåt:er med lang å og dobbel t
Slup	Klut	Slu:p med lang u
Skag	Skog	Ska:g med lang a
Kjæst	Kjær	Çæ:st med lang æ Bokstavkombinasjonen <kj> blir uttalt ç, som i kjær
Glunk	Dunk, bank	Glʊŋ:k med trykk på ŋ Bokstaven <u> foran <nk> uttales ʊ, som i munk, dunk. Bokstavkombinasjonen <nk> blir uttalt ŋ:k, som i bank.
Freip	Fleip	Fræɪ:p med trykk på æɪ Bokstavkombinasjonen <ei> blir uttalt æɪ, som i fleip eller stein.
Tjesk	Tjern eller Tjene	Çes:k med trykk på s Bokstavkombinasjonen <tj> blir uttalt ç, som i tjern. Unntak er tjener, tjene og tjeneste. Non-ordet «Tjesk» kan derfor uttales på to ulike måter.
Grast	Brast	Gras:t med trykk på s
Frybe	Rype	Fry:be med lang y
Spurl	Spurv	Spu:rl med lang u
Blimp	Blink	Blim:p med trykk på m
Kyrtos	Kyst eller Kypros	Bokstavkombinasjonen <ky> blir uttalt çʏ, som i kyss, kyst.

		Bokstavkombinasjonen <rt> blir uttalt ʈ, som i hjerte
Skjape	Skjære	ʃa:pe med lang a Bokstavkombinasjonen <skj> blir uttalt ʃ, som i skjev
Bedper	Bed (Med) Per (Ser)	Be:dpe:r med lang e. Bokstaven <d> blir ofte stum i slutten av ord etter vokal og når man legger til endelse, slik som for eksempel ved, glad, glade osv. I dette tilfellet er bokstaven <d> i ordet, så her blir ikke <d> stum.
Jittop	Jitt (Gitt) Op (Opp)	Jit:åp: med å og dobbel t og p
Flåmbre	Flåm (Flom) Bre (Bred)	Flå:mbre med lang å
Plyktus	Kaktus	Plyktus: med trykk på s
Kæpken	Kæp (Kapp) Ken (Men)	Kæp:ken med trykk på p
Nalplir	Nal (Nål) Plir (Blir)	Na:lplir med lang a
Segvok	Seg (Legg) Vok (Bok)	Seg:vok med trykk på g
Dronperk	Dron (Krone) Perk (Lerke)	Dro:npærk med æ og lang o
Flønders	Flønd (Flom) Ers (Vers)	Fløn:deʃ med trykk på n Bokstavkombinasjonen <rs> blir uttalt ʃ, som i mars.
Kjændelsk	Kjæn (Kjær) Delsk (Falsk)	Kjæn:delsk med trykk på n
Pramfert	Pram (Fram) Fert (Fart)	Pram:feʈ: med trykk på m og ʈ Bokstavkombinasjonen <rt> blir uttalt ʈ, som i ordet hjerte
Tjemkrup	Tjem (Tjern) Krup (Kryp)	Tje:mkru:p med lang e og u
Stromklått	Strom (Strøm) Klått (Flott)	Strom:klåt: med trykk på m og dobbel t
Høfignak	Høfig (Høflig) Nak (Sak)	Høf:ina:k med trykk på f og lang a
Terlingdom	Terling (Erling) Dom (Rom)	Tæ:liŋ:dom med lang æ og trykk på ŋ: Bokstaven <e> foran <rl> blir uttalt æ, som i navnet Erling. Bokstavkombinasjonen <ng> blir uttalt ŋ: som i ordet mange.
Satrekrant	Satre (Sadle) Krant (Brant)	Sa:trekrant med lang a
Ulterpolm	Ulter (Sulter) Polm (Holm)	Kan uttales ul:terpål:m eller ul:terpol:m
Eboterlart	Ebo (Demo) Ter (Ler) Lart (Fart)	Ebo:tela:ʈ Bokstavkombinasjonen <rl> blir uttalt l, som i perle. Bokstavkombinasjonen <rt> blir uttalt ʈ, som i ordet hjerte.

Nonword B

Non-ord	Reelle ord, eksempel	Østnorsk uttale/regler
Ma	Sa	M:a med lang a
Py	Fy	P:y med lang y
Nu	Nå	N:u med lang u
Ir	Yr	I:r med lang i
Eb	Rebus	E:b med lang e
Ot	Sot	O:t med lang o
Øb	Løpe	Ø:b med lang ø
Rin	Lim	Ri:n med lang i
Dutt	Kutt	Dut: med dobbel t
Kas	Mas	Ka:s med lang a
Dav	Lav	Da:v med lang a
Bev	Vev	Be:v med lang e
Sod	Bod	So:d med lang o
Fyt	Nyt	Fy:t med lang y
Hame	Dame	Ha:me med lang a
Dære	Pære	Dæ:re med lang æ
Ferg	Ferge	Ferg med <æ>, som i ferge
Huss	Nuss	
Kjar	Kjær	Bokstavkombinasjonen <kj> blir uttalt ç, som i kjær.
Snåp	Snop	
Nukke	Dukke	
Parb	Park	
Krip	Grip	
Tvip	Skvip	
Flem	Flåm	
Polk	Folk	Polk med å, som i folk
Væst	Vest	
Mosse	Tasse	
Skjaps	Skjære	Bokstavkombinasjonen <skj> blir uttalt ʃ, som i skjev
Nuppi	Nuppe	
Skel	Skål eller ski	Bokstavkombinasjonen «sk» blir uttalt –sk eller ʃ
Fyli	Falig	
Slert	Stjert	Bokstaven <e> foran <rl> blir uttalt æ, som i navnet Erling og bokstavkombinasjonen <rt> blir uttalt ʈ, som i ordet hjerte.
Spife	Spise	
Grol	Sol	

Preit	Greit	Bokstavkombinasjonen <ei> blir uttalt æɪ, som i fleip eller stein.
Hjikk	Gikk	Bokstaven <h> foran <j> blir stum.
Kjøft	Kjeft	Bokstavkombinasjonen <kj> blir uttalt ç, som i kjær.
Grysk	Lyst	
Åtinn	Å (Å) Tinn (Tann)	
Faldig	Farlig	Faldig uttales med stum <g>
Lutesk	Grotesk	
Trakki	Tråkke	
Krøber	Kryper	
Belote	Be (Be) Lote (Mote)	
Druker	Dråper	
Tigler	Tiger	
Utlevig	Utvendig	Utlevig uttales med stum <g>
Skrykent	Strøkent	
Hyglert	Hyg (Byg) Lert (Hjerte)	Hyglert uttales med lang <y> og bokstavkombinasjonen <rt> blir uttalt ʈ, som i ordet hjerte.
Brumisk	Komisk	
Vonend	Vone (Sone) Nd (Nd)	
Kradon	Radon	
Bløkent	Strøkent	
Flinpikt	Flin (Fin) Pikt (Plikt)	
Slopnårt	Slop (Snop) Nårt (Sårt)	Bokstavkombinasjonen <rt> blir uttalt ʈ, som i ordet hjerte
Skrøtter	Skvetter	
Lypremisk	Lypre (Hypre) Misk (Fisk)	
Gnunærte	Gnun (Snu) Ærte (Erte)	
Masplinakt	Mas (Mas) Plin (Fin) Akt (Akt)	
Svyrtuknot	Svyr (Svir) Tuk (Luke) Not (Note)	
Elpandorikk	Elektronikk	
Apotraknatet	Apo (Ape) Traknatet (Attentatet)	Apotraknatet uttales med stum <t> i slutten av ordet