Receptive vocabulary in oral text reading fluency

A quantitative study on the relationship between receptive vocabulary and oral text reading fluency in grade 3 children.

Katrine Thams Leonard

Master thesis in special needs education
Department of special needs education
University of Oslo

SPRING 2020

Abstract

Reading fluency is fast and accurate word recognition and once established, closely linked to comprehension. Fluency is thought to free up cognitive capacity allowing for integration of words in a text and its development is critical to be considered a skilled reader. Academic and professional success is reliant on skilled reading, hence the importance for the educational field to understand the processes behind development of adequate reading skills.

The purpose of the present cross-sectional study is to determine to what extent receptive vocabulary can predict oral text reading fluency in grade 3 children. Through hierarchical regression analysis, age and skills thought to capture word reading in sequence are used as control variables in the hope of identifying the unique variance explained in oral text reading fluency by receptive vocabulary. The sample consists of 70 Dutch grade 3 children who have all been assessed on oral text reading fluency, receptive vocabulary, individual word reading, serial word reading rate (word lists) and serial digit naming.

The results show that receptive vocabulary does not explain any statistically significant unique variance in oral text reading fluency in a relatively transparent language in grade 3. The study challenges the idea that established readers rely on elements of vocabulary for text reading fluency. However, the large amount of unexplained variance after controlling for elements thought to capture reading of words in sequence should not be overlooked.

The data was collected in a relatively transparent orthography where accurate decoding skills are thought to become established earlier than in opaque orthographies. Thus, the assumption that a language skill like receptive vocabulary would explain unique variance in text reading fluency is not unreasonable, but the results are somewhat surprising.

Preface

Over the years and through practical experience in schools I have developed a particular interest in early reading development. Thus, the completion of the Master in special needs education was a good opportunity to explore the complexities of research conducted in exactly this field. Going forward, I take with me the importance of critical thinking and of sustaining an analytical and systematic approach not only to research, but also to assessments and remediation in education.

Thank you to my supervisor and lecturer Athanassios Protopapas for relaying some of his vast knowledge in the field and for being available to give advice throughout this somewhat challenging period due to Covid-19.

Thank you also to Sietske van Viersen for advice, background information and access to data from the study on orthographic learning in the Netherlands which has been analysed in the present study.

And finally, thank you to my family for their loving support throughout.

Katrine Thams Leonard
Oslo, 29.06.2020

Contents

Word count: 17434

List of figures

List of tables

1. Introduction

Reading is considered a core skill in the Norwegian school system and is imperative to master for academic and professional success (Utdanningsdirektoratet, 2017). Skilled reading develops through practice and integration of subskills involved in individual word recognition (Ehri, 2005). However, skilled reading requires abilities beyond effiecient word recognition, abilities thought to be related to language comprehension. Fluency is fast and efficient integration of words allowing for comprehension processes. Thus, reading fluency is a good indicator of overall reading competence and an essential part of skilled reading (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003; Kuhn, Schwanenflugel, Meisinger, Levy, & Rasinski, 2010).

Once word decoding becomes faster and texts more demanding, comprehension relies on other factors like language comprehension (Catts, 2018). This shift seems to happen around third to fourth grade in English and even earlier in more transparent languages (Tilstra, McMaster, Broek, Kendeou, & Rapp, 2009; Verhoeven, van Leeuwe, & Vermeer, 2011).
It seems that more established readers no longer only rely on accurate word decoding for comprehension. This however varies depending on the transparency of the language, but it seems other skills than decoding affects fluency in context in transparent and established readers (Lervåg, Hulme, & Melby-Lervåg, 2018; Protopapas, Simos, Sideridis, & Mouzaki, 2012; Rakhlin, Mourgues, Cardoso-Martins, Kornev, & Grigorenko, 2019). Research on the topic of vocabulary and reading fluency is scarce, and few studies break down vocabulary to smaller aspects of language and connect them to fluency or comprehension (Ouellette, 2006; see also Braze et al., 2016; Language and Reading Research Consortium, 2015; Park & Uno, 2015; Protopapas, Mouzaki, Sideridis, Kotsolakou, & Simos, 2013; Tilstra et al., 2009; Verhoeven et al., 2011).

Fluency is found to explain unique variance in comprehension and comprehension affects fluency suggesting that the semantic properties affecting comprehension also affects fluency (Jenkins et al., 2003; Tilstra et al., 2009). A reciprocal relationship between vocabulary and comprehension has also been established (Verhoeven et al., 2011), although vocabulary has been found to have less of an influence on reading ability in older readers in a transparent language (Carretti, Toffalini, Saponaro, Viola, & Cornoldi, 2019).

The scarcity of information on the connection between language skills and oral reading fluency is regrettable as more knowledge in the area could have important implications for both educational assesssment and practices.

The present study seeks to address this gap and build on existing literature by analysing data on receptive vocabulary and oral text reading fluency from a relatively transparent orthography (Dutch) in grade 3 where children's decoding skills are expected to be established.

The following research question is posed:

To what extent does receptive vocabulary knowledge predict oral text reading fluency in Dutch grade 3 children?

To answer the research question, the study will examine whether receptive vocabulary can uniquely explain variance in oral text reading fluency, once age and elements thought to capture reading of word sequences have been controlled for.

1.1 Clarification of concepts

Word reading, decoding, accurate word reading, and individual word reading are used interchangeably, but relates to the processes of reading single words. Discrete word reading relates to the format in which word reading is measured in assessment.

Serial word reading relates to reading of more than one word and is also referred to as reading words in sequence. It can take place in a list format or in a continuous text (context).

Connected text relates to a passage and means reading of words in context rather than individually.

Language comprehension is a wide term and in the present study is related to aspects of oral language like listening comprehension, semantic[1] knowledge and vocabulary.

---

[1] Semantic is a linguistic term and related to building word meanings and connections in vocabulary (Sveen, 2011).

1.2 Structure

Chapter 2 is the literature review where oral reading fluency is defined. An explanation of its importance in reading development is provided. The chapter addresses individual word reading, text level processing, language skills, text reading and comprehension, serial processing skills and finally links oral reading fluency and vocabulary.

Chapter 3 describes the method applied in the study and considers ethical dilemmas.

Chapter 4 outlines the descriptive statistics, bivariate correlation and hierarchical regression analysis and the results found.

Chapter 5 is a discussion considering both theoretical findings and validity and reliability. The chapter also has a section on limitations and implications for education and future research.

Chapter 6 is the conclusion.

2. Literature Review

Fluent reading of text is the coordination of underlying component skills, leading to accurate and efficient word recognition in sequence which allows for comprehension (Fuchs et al., 2001). Fluency is an essential part of skilled reading (National Reading Panel, 2000). The literature review will define concepts and address relevant theory regarding the development of oral reading fluency (ORF). Further the component skills underlying skilled reading will be addressed as they are related to the actual development of fluency, hence the importance to consider the relationship between these components and their connection to ORF. Vocabulary explains unique variance in decoding and comprehension and is consequently linked to skilled reading. Vocabulary constitutes both words you perceive and words you produce and in the present study the focus is on words perceived, namely receptive vocabulary (Ouellette, 2006).

2.1 Oral reading fluency

The goal of reading instruction is to enhance fluency and thereby facilitate reading comprehension. Reading fluency can be considered a bridge between decoding and comprehension and is the ability to read a text «at speed, accurately and with proper expression» (National Reading Panel, 2000, p. 3-1; Pikulski & Chard, 2005). ORF is a distinct feature of skilled reading where the reader recognises words rapidly and maintains speed whilst moving through a text. Accurate and fast word recognition is fundamental in fluency and to allow for comprehension processes. Comprehension is considered a complicated cognitive process integrating word reading and language skills and is necessary for skilled reading (Fuchs et al., 2001; National Reading Panel, 2000; Pikulski & Chard, 2005). Reading with proper expression (or prosody) means applying the rhythm and the intonations of the language. This is also important for comprehension. Prosody is rarely a part of fluency measures (Kuhn et al., 2010) and is therefore beyond the scope of this study.

Over the past two decades there has been an increased importance of ORF in the literacy curriculum and in reading research, possibly since the National Reading Panel (2000) views fluency as one of five critical components of reading development. In both research and classroom assessment, ORF is often a timed measure of accurate decoding of wordlists or words in a text, but text fluency has been found to uniquely predict more variance in comprehension than word lists (Fuchs et al., 2001; Jenkins et al., 2003). Reading typically occurs in a connected text format rather than lists and the strong connection between ORF and comprehension highlights the importance for the field to understand the underlying processes and skills important for oral text reading fluency.

2.2 Components of skilled reading

Skilled reading is intricate, it is both simple and complex. Simple in that once mastered the process seems effortless and autonomous and complex in that it does not just develop naturally but requires practice. Development of underlying skills is necessary for word reading. These skills are thought to be retrieval of letter sounds, phonological awareness, blending and decoding (Ehri, 2005; Fuchs et al., 2001; Logan, 1997). The focus in the present study lies beyond acquisition of these underlying skills, examining what affects ORF in terms of it being an indicator of skilled reading. This chapter addresses individual word reading and the theory of automaticity followed by text-level processing. Finally, it touches upon the role of language skills in text-level processing and comprehension.

2.2.1 The simple view of reading

There is a general consensus amongst researchers that fluency in word reading is related to the ability to comprehend the text that is being read and that reading for understanding is the goal of reading instruction. Essential components of reading can be understood through the simple view of reading (the simple view; Gough & Tunmer, 1986; Hoover & Gough, 1990; Jenkins et al., 2003; Perfetti, 1985). The framework posits that reading comprehension (R) is the product of two broad components, namely decoding (D) and language comprehension (L), ($R = D \times L$). The simple view is not a model for reading development, but merely a framework for conceptualizing reading as being comprised of two distinct components.

Testing the predictions of the simple view has been challenging for researchers as the components are broad and therefore often defined differently across various studies. However language factors (grammar, receptive and expressive vocabulary, verbal working memory and inference skills) combined with decoding skills have been found to explain almost all the variance in reading comprehension in grade 2 when examining Norwegian children (Lervåg et al., 2018) and similar findings occur for American children (Tilstra et al., 2009) in grade 4 when examining decoding and listening comprehension. In terms of the decoding component specifically Protopapas et al. (2012) found in their study of Greek children that word and nonword reading can be combined and load on the same subskills, but accuracy (decoding) and fluency (timed measures of word reading) are distinct elements and cannot be measured together, also causing complexity in terms of how the decoding construct in the simple view should be measured. However, the studies demonstrate a strong relationship between decoding, language comprehension and reading comprehension.

The article “The simple view of reading” by Hoover and Gough (1990) has been quoted more than 2900 times according to Google Scholar. As a well-established framework and a source of numerous studies it has also been used in the present study to create a link between comprehension, decoding, language skills and ORF. However, the intention of the present study is not to evaluate the separate components’ contribution in the framework.

2.2.2 Individual word reading

Word reading (decoding skill) is necessary for skilled reading and explains significant variance in reading ability particularly in early development (Hoover & Gough, 1990). Another recognized framework as a premise for understanding the development of word reading is Ehri’s (2005) phase-theory. It characterizes reading development as a process moving through several phases where recognising words rapidly from memory is the goal and foundation in skilled reading. The reader progresses through phases from becoming familiar with the alphabetic principle, establishing grapheme-phoneme correspondence, blending of letters, and eventually becoming a sight word reader. A sight word reader is one who from memory automatically recognises words and their spelling. It requires practice to establish fast and accurate representation of words and to enhance sight word vocabulary in memory (lexicon; Perfetti, 2007). Development of sight word reading is thought to also take place in transparent orthographies[2], although Ehri’s (1998, 2005) research has been conducted in English. The properties of the phases and their significance in reading development may be slightly different (Ehri, 1998, 2005; Perfetti, 2007).

---

[2] A transparent orthography is one that has consistent mappings between letters and phonemes whereas an opaque orthography is inconsistent. A large body of research has been conducted on the development of reading in English whereas Dutch (and Norwegian) are relatively transparent orthographies (Seymour, Aro, & Erskine, 2003). Throughout this study it has therefore been important to bear in mind that children become fluent readers later in English than in Dutch (and other transparent languages) and be aware of this significant difference when generalizing from other research conducted in English (Ehri, 2005).

By acquiring a large sight word vocabulary, children process individual words more efficiently rather than going through laborious decoding processes every time they read a word. Although good decoding skills contribute to sight word reading (Rakhlin et al., 2019) it is not the only factor that influences the speed at which words are recognised. It is thought that automaticity in word recognition is a process that helps free up cognitive capacity allowing for comprehension (Logan, 1997). If any of the processes involved in reading lack efficiency a reader will use too many resources and thereby hamper the comprehension process. Prior to the mid-70s, research on automatic cognitive processes had primarily focused on the development of automatic perceptual-motor skills but from the mid-70s reading as an automatic process also became an area of interest. Automaticity in reading was acknowledged as an important component of skilled reading and is today thought to comprise of four main properties: speed, effortlessness, autonomy, and lack of conscious awareness (Kuhn et al., 2010; Logan, 1997). For beginner readers, word reading is slow and effortful, whereas skilled reading is automatic and described as fast and effortless. With practice the connections between word and letter patterns are thought to be strengthened and stored in long-term memory, thereby increasing automaticity of retrieval. Reading speed is thus increased and becomes fast in automatic processing, also referred to as the instance theory of automatization. When little attention is given to one task (decoding), it gives the reader the opportunity to give attention to other tasks (comprehension; LaBerge & Samuels, 1974; Logan, 1997). Skilled reading is also autonomous and happens without consciousness and intention. A skilled reader automatically recognises words without giving the process any special attention whereas for a beginner reader decoding is laborious (Logan, 1997).

The relationship between sight word reading and fluency has recently been demonstrated in Russian which like Dutch is a relatively transparent language. Rakhlin et al. (2019) found that sight word reading (measured with word lists) was the strongest correlate with reading fluency over and above that of word decoding accuracy in grade 3 children ($r = 0.535$). The study also showed that sight word reading significantly predicted oral reading fluency in good readers but neither individual word reading nor nonword reading had the same effect. For poor readers, this relationship was not present as they are still relying on basic decoding strategies indicating the importance of decoding skills in early reading development. The study highlighted the importance of efficient sight word reading in fluency.

As mentioned above there is not one obvious and acknowledged way to measure the components of the simple view. Word reading is often measured as accurate word reading of single words or nonwords, but there is no clear consensus as to whether word reading should be a pure decoding (accuracy) measure or a combined accuracy/speed (fluency) measure. This lends confusion to the field and possibly partly explains the range of results in variance explained by word reading in comprehension (García & Cain, 2014).

The Language and Reading Research Consortium (2015) examined whether accuracy (decoding) and fluency made separable contributions to comprehension in a cross-sectional analysis. Their sample consisted of English-speaking children in grades 1, 2 and 3. They found that fluency was a separate construct in comprehension, especially in later years when reading skills were more established. Children are more dependent on accurate decoding skills in early reading development supported by findings in the Rakhlin et al. (2019) and Protopapas et al. (2012) studies mentioned above. In addition, both Jenkins et al. (2003) and Tilstra et al. (2009) reported strong relations between fluency measures used in their studies and reading comprehension. It seems fluency is a better predictor of overall reading competence (assessed with tests of reading comprehension) in the two studies than word list and accurate decoding of nonwords (respectively). This particularly applies in later years when decoding is more established, suggesting fluency should also be measured as part of overall reading ability and to understand differences in comprehension when decoding skills are established (Catts, 2018; Language and Reading Research Consortium, 2015; Protopapas et al., 2012; Tilstra et al., 2009).

2.2.3 Text-level processing

So far, the importance of the development of accurate individual word reading in connection with fluency and comprehension has been highlighted, mainly emphasizing phonology as the important factor in word reading (Ehri, 1998, 2005). In addition, the theory on automaticity illustrates how word recognition needs to be both fast and accurate for fluency to develop (LaBerge & Samuels, 1974; Logan, 1997). Reading fluency is often measured in either list format or in a continuous text, although it seems reading in context facilitates reading speed of oral reading, at least for more proficient readers and also mimics “actual” reading which usually takes place in context (Jenkins et al., 2003).

The interactive model of reading (Stanovich, 2000) expands on the theory of automaticity mentioned above (LaBerge & Samuels, 1974), incorporating influence from text as another factor affecting fluency. The model assumes that interaction with words in context affects processing speed and aids comprehension processes. It follows that text processing is different from individual word processing in that language skills facilitate both fluency and comprehension (Hoover & Gough, 1990; Jenkins et al., 2003). Therefore, reading fluently in context should be a good indicator of overall reading competence as found by Fuchs et al. (2001) who reported a high correlation between ORF measures and comprehension.

The concept of reading in and out of context was examined by Jenkins et. al (2003) in a cross-sectional study. They examined English speaking children in grade 4 on context-free (word reading measured with lists) and context reading fluency (measured with passages). They found that comprehension was better accounted for by context fluency (number of correctly read words per second) than by list fluency (29% vs 4% respectively). In addition the relationship was reciprocal with reading comprehension skill uniquely predicting context fluency indicating that the more you understand the faster you move through the text and that the semantic representations (language abilities) that affect comprehension also affects ORF, an important finding for the present study.

Jenkins et. al. (2003) demonstrated that fluent reading of text is strengthened through semantic activation of words which is linked to the language component of the simple view. However, Schwanenflugel et al. (2006) did not find in their cross-sectional study of English-speaking children in grade 1 to 3 that text-reading fluency predicted any additional variance in comprehension beyond word reading fluency. Although these findings contradict Jenkins et al. they may merely represent a developmental perspective as the participants in Schwanenflugel et al. study were younger and less established readers. Their sample also consisted of a larger proportion of children receiving subsidized lunch which may have an indirect implication in terms of their literacy level. In the U.S children with lower socioeconomic status (SES) tend to have a lower literacy level and only low income families receive free or subsidized lunch (Byrnes & Wasik, 2019). Another factor which may explain this discrepancy is that the texts used for comprehension were not demanding enough in early years of schooling (grade 1 to 3). The findings in Schwanenflugel et al. do however support the role of decoding in early reading development and are in line with a recent cross-sectional study by Altani et al. (2019). They found reading of individual words and word lists to strongly correlate in Greek children in grade 1 (r = .84), but only moderately in grade 5 (r = .56). The above results indicate the significant role of decoding for early reading development and that other factors related to language skills and sequential processing play a role when reading becomes more established.

Both accuracy and speed are important for oral text reading fluency, although weaker readers seem to rely more on accuracy than more experienced readers who rely more on language skills. Reading in context seems to facilitate speed of reading which also increases with school age suggesting that as decoding is established, speed can pick up, thus suggesting that reading words in context is faster and facilitates comprehension processes and vice versa (Jenkins et al., 2003).

2.2.4 Language skills in text level processing and comprehension

As highlighted above comprehension involves processes beyond individual word recognition and these are typically thought to be related to language comprehension as per the simple view (Hoover & Gough, 1990). A range of instruments are used to assess language skills across different studies. These could be oral versus silent tasks, a variety of vocabulary measures, listening comprehension, grammar and test of inference skills (Braze et al., 2016; Language and Reading Research Consortium, 2017; Ouellette, 2006). This adds complexity to the field. Oral vocabulary measures have been found to be good indicators of overall language comprehension (Protopapas et al., 2012) and listening and language comprehension have been found to be highly correlated (Language and Reading Research Consortium, 2017). Vocabulary is part of a child's language skills and has been found to account for unique reading comprehension variance after controlling for listening comprehension, though not omitting other aspects of language which may be part of the language component of the simple view (Braze et al., 2016).

Vocabulary underlies the ability to understand oral language and to comprehend text when reading. Unknown words can cause difficulty in inferring meaning in context, but by acquiring new words (more semantic entries) children and adults alike build their mental lexicon (Perfetti, 2007). Complete knowledge of a word includes an array of linguistic knowledge like knowing the words spelling, pronunciation and syntactic and semantic relation to other words. A child's vocabulary range can be split into receptive and expressive vocabulary, meaning the words understood when heard or read and the words used orally or in writing. Children and adults alike will generally have more words they understand than words they use in their vocabulary (Nation, 1990). The development of both receptive and expressive vocabulary tends to occur together and they both depend on the other for development (Burger & Chong, 2011). In addition, the literature also refers to breadth as the sheer quantity of words known and depth of vocabulary as having detailed knowledge of the word's semantic properties (Ouellette, 2006).

Experience with words facilitates growth of vocabulary and development of semantic representations. Text reading fluency is characterized by efficiently integrating the orthographic and semantic knowledge of words and thereby freeing up capacity to consolidate the information in the text (Fuchs et al., 2001). Lexical quality theory emphasizes the importance not only of orthographic representation, but also knowledge of words' semantic properties. High quality lexical representations make words easier to recognize when reading in context. Having too many low quality representations of words would hamper the comprehension process. Not only do children need high quality representations in their lexicon, they also need to access these words efficiently as per the verbal efficiency theory (Perfetti, 1985). Efficient access frees up capacity in verbal working memory allowing for comprehension. Both lexical access and semantic encoding must be efficient otherwise oral reading fluency would be affected. Limited access to words' meaning will hamper oral reading fluency and thereby affect skilled reading (Perfetti, 1985, 2007). The lexical quality theory relies more on quality of single word representations than the interaction between decoding and language skill as per the simple view.

The role of vocabulary in reading is not yet fully understood by researchers, although it has been recognised as an essential component for developing reading comprehension by the National Reading Panel (2000). Vocabulary seems to account for unique reading comprehension variance after controlling for oral language skills both in Canadian and Greek children (Ouellette, 2006; Protopapas et al., 2013). But vocabulary has also been found to directly affect comprehension in American children through both decoding and language comprehension as per the simple view (Language and Reading Research Consortium, 2015). This complicates assessment of the various components and possibly highlights the importance of a developmental perspective on the simple view, emphasizing decoding as an indicator of reading skills in the early years of schooling and thorough assessment of language skills later (Catts, 2018). This shift in the dominance of language skills when decoding skills become faster and more established seems to happen around grade 2 to 4 depending on the transparency of the language. Children gradually become more reliant on language factors like vocabulary when the texts they read become more demanding (Braze et al., 2016; Tilstra et al., 2009). However, Protopapas et al. (2013) found that fluency did not account for any unique variance in comprehension in Greek speaking children (grade 2 – 5) neither concurrently nor longitudinally although accuracy seemed to account for a significant amount of unique variance in comprehension concurrently, concluding that accuracy measures also have importance in transparent languages for intermediate readers as long as they are complex enough.

The developmental perspective on vocabulary and reading skills have also been addressed in a Dutch study. In the longitudinal study from grade 1 to 5, children's vocabulary skills early on in reading instruction (grade 1) affected early decoding skills and comprehension. There was also a strong association between early vocabulary skills and comprehension later. Reciprocally, decoding promoted vocabulary development in intermediate readers (grade 2 to 5), indicating that efficient word decoding is a key to vocabulary growth. A reciprocal relationship was also found for comprehension and vocabulary growth in lower grades (1-3). Developmentally it seems knowledge of words in a text can facilitate word decoding and comprehension of the text and that skilled reading of text stimulates growth of vocabulary. The findings support the lexical quality hypothesis where quality of word representations support development of comprehension and vice versa (Verhoeven et al., 2011). In the Greek study mentioned above (Protopapas et al., 2013) listening comprehension and receptive vocabulary contributed unique variance to reading comprehension longitudinally and also concurrently. In addition to there was a strong association between reading accuracy and comprehension. There was however a challenge in separating the two constructs (vocabulary and listening comprehension) causing difficulties in interpretation in terms of the simple view. A possible explanation is related to construct validity and how listening comprehension versus vocabulary were measured. The findings are however more in line with the lexical quality hypothesis where vocabulary is an index for overall lexical skills in terms of knowledge of words' orthographic and semantic properties.

The developmental perspective is supported in other studies. Tilstra et al. (2009) found in a study of American children and carried out across grade levels (grade 4, 7 and 9) that the variance accounted for in comprehension by decoding and listening comprehension measures decreased with grade levels (61% to 38%). Listening (language) comprehension accounted for an increasing amount from grade 4 to 7 (19% to 35% respectively) with decoding decreasing from 42% to 13%, indicating the increased importance of language skills as decoding is established and efficient. Similarly, it was found in Norwegian children (grade 1-6) that for poor decoders (and early years of schooling) differences in reading comprehension varied more with differences in decoding skills whilst good decoders' comprehension benefitted more from language skills. Meaning that the nature of the relation between decoding and comprehension changes as the reader becomes more advanced, suggesting that reading fluency may share more variance with listening comprehension when reading is more established (Lervåg et al., 2018).

Somewhat counterintuitively, Carretti et al. (2019) found in a transparent orthography that the proportion of variance explained in text reading speed by reading comprehension decreased with school age. Silent and oral comprehension measures were used to describe the construct. They examined Italian children in grades 3-5, 6- 7 and 9-10. The transparency of the language may explain the discrepancy in these findings, although does not necessarily explain the different results from Verhoeven et al. (2011) which also examined a transparent orthography, although the children in Carretti et al. were significantly older.

Disentangling the direct impact of certain language abilities on reading skills are important in the present study which addresses receptive vocabulary. In a Canadian study (Ouellette, 2006) receptive vocabulary was considered an indicator of a child's breadth of perceptive vocabulary skills and expressive vocabulary an indicator of breadth of productive vocabulary, whereas depth of vocabulary was represented by semantic knowledge (measured with word definitions and synonyms). Interestingly, receptive vocabulary knowledge was found to explain unique variance in sight word reading (word list) when decoding skills (nonword) were controlled for although subsumed by expressive vocabulary when alternating the entry in a regression analysis. Semantic knowledge explained unique variance in sight word reading when entered first in a regression analysis (3%) followed by expressive vocabulary (3.3%) which subsumed receptive vocabulary. Results indicate that both semantic knowledge and vocabulary breadth influence sight word reading, although the direct impact seems to get subsumed by other aspects of vocabulary. Unique variance was also explained by semantic knowledge in comprehension (12.1%), subsuming both receptive and expressive vocabulary (vocabulary breadth). The findings show that semantic knowledge is the prime indicator of comprehension, whereas breadth of vocabulary knowledge has more of an impact on sight word reading.

Ouellette (2006) points out that studies measuring semantic knowledge have a stronger relation to reading comprehension than those specifically examining vocabulary breadth. Yet as both sight word reading and comprehension seem to be affected by vocabulary measures it would be reasonable to expect reading fluency, as the bridge between the two, to also be affected by vocabulary skills. This is supported by a cross-sectional study of Hangul[3] (grade 1 to 4) where receptive vocabulary was found to predict significant unique variance in text reading fluency in grade 2. The study also found receptive vocabulary to be a predicter of decoding (nonwords). The study looked at the contribution of various cognitive abilities (visual cognition, phonological awareness, naming speed and receptive vocabulary) on reading and spelling and also highlighted the discrepancies between different orthographies. English seems to rely on accurate decoding skills for comprehension for longer than more transparent orthographies (Park & Uno, 2015).

It seems language abilities affect decoding, fluency, and comprehension and that being exposed to more words affects your vocabulary skills. In addition, the reciprocal relationship between vocabulary and comprehension and fluency and comprehension means comprehension also affects speed of processing and vocabulary growth, at least for younger readers (Carretti et al., 2019; Jenkins et al., 2003; Verhoeven et al., 2011).

2.3 Serial processing of words

Skilled reading of connected text is not just integration of word recognition and language comprehension, but also involves moving through a series of words in a row. The sequential processing needs to happen at a certain speed to be efficient. Serial word reading involves parafoveal previewing of the words allowing for comprehension of a text before pronouncing the actual words. Serial processing of words has been shown to be connected to rapid automatized naming (RAN; Protopapas, Katopodi, Altani, & Georgiou, 2018).
RAN refers to the time "required for a child to quickly and accurately name an array of well-known visual stimuli (usually letters, digits, objects, or colours)" (Araújo, Reis, Petersson, &

[3] Hangul is the Korean phonetic alphabet. Although syllabically structured, it is still fundamentally based on an alphabetic structure with 14 consonants and 12 vowels. Hangul is considered a transparent orthography.

Faísca, 2015, p.868). A meta-analysis conducted by Araujo et. al. (2015) investigated the strength between RAN and reading and identified variables mediating the relationship by examining a large body of research. They found a moderate to strong correlation between RAN and reading ability (r = 0.43) suggesting that RAN taps into cognitive processes also involved in reading. The coefficients were higher for text reading than nonword reading suggesting a stronger link to serial word reading than single word reading. The meta-analysis also found letter- and digit- naming tasks to be more strongly related to reading competence than naming of colours and objects (Araújo et al., 2015). This is supported in a recent study by Altani et al. (2019) who found serial digit naming to be a significant and unique predictor of both serial word reading (word lists) and text reading fluency for intermediate readers (Greek children in grade 3 and 5). The cross-sectional study showed that the correlation between single word reading and word lists and connected text reading decreases over the course of development meaning individual and serial word reading are distinct constructs in more advanced readers. Word list reading fluency became increasingly more like text reading fluency than individual word reading, indicating that another processing skill beyond language may have an influence on serial word reading as language processing skills are not dominant in word lists. RAN is thus thought to represent a processing factor reflecting the difference between individual word reading and word lists/text reading. It seems that RAN measures (digit naming) capture the processes that distinguish individual word reading from serial word reading (at least for word lists as the text used in the study was less demanding).

2.4. Oral text reading fluency and vocabulary

Fuchs et al. (2001) proposed in their analysis that ORF entails the ability to "process meaningful connections within and between sentences, relating text meaning to prior information, and making inferences to supply missing information" (p. 240). ORF is also defined as the bridge between decoding and reading comprehension (Pikulski & Chard, 2005). Studies investigating the direct relationship between receptive vocabulary and oral text reading fluency are scarce. Apart from Park and Uno, (2015) no studies have been identified that specifically look at the combined role of receptive vocabulary skills and reading fluency in texts, although several studies have looked at language skills and used decoding, sight word reading or comprehension as their outcome and found a connection (Braze et al., 2016; Hoover & Gough, 1990; Jenkins et al., 2003; Ouellette, 2006; Tilstra et al., 2009; Verhoeven et al., 2011).

Receptive vocabulary has been found to be a unique predictor of decoding, and vocabulary knowledge in general has been found to be a strong predictor of reading ability when controlling for decoding (Ouellette, 2006). Research also indicates that reading comprehension predicts connected text fluency after controlling for individual word reading rate. This finding suggests that the semantic knowledge that affects comprehension also affects oral reading fluency (Jenkins et al., 2003) at the same time as there is evidence that fluency constitutes fast and efficient recognition of sight words (Rakhlin et al., 2019). Reading fluency of a connected text explained additional variance (8%) over and above that of word-level decoding, listening comprehension and verbal proficiency (language comprehension) in Tilstra et al. (2009). The finding suggests that fluency is a separate construct in comprehension beyond that of decoding, listening comprehension and vocabulary knowledge. ORF seems to be a separate predicter of reading competence suggesting that ORF can be used as a proxy for comprehension (Fuchs et al., 2001; Tilstra et al., 2009). ORF has been found to have a significant positive relation to reading proficiency; sight word reading and comprehension, but with a changing role throughout development (Jenkins et al., 2003; Rakhlin et al., 2019; Verhoeven et al., 2011).

The above theoretical findings claim that skilled reading is a consolidation of efficient word recognition in context, influenced by language skills and the ability to process words in sequence. Measures of receptive vocabulary can capture much of the relevant language skills variance, whereas individual word reading, serial word reading, and RAN should theoretically capture the processes underlying word reading in sequence.

Thus, based on the above theoretical findings it is hypothesized that there should be a relationship between receptive vocabulary and oral text reading fluency. In particular, it is hypothesized that receptive vocabulary, as a proxy for language skills will explain unique variance in oral text reading fluency after controlling for elements thought to capture reading of word sequences (individual word reading, serial word reading and RAN).

3. Method

This chapter will describe the research method applied in the study, including its design, procedure, participants, and assessment tools. Further there will be a section on validity and reliability and finally a section on ethical considerations.

3.1 Design

The purpose of the present study was to investigate the relationship between receptive vocabulary and oral text reading fluency. By examining data borrowed from a large study on orthographic learning in the Netherlands I sought to predict the contribution of receptive vocabulary on oral text reading fluency in grade 3 children, after controlling for age and elements of word reading in sequence. The sample was picked from four schools in both urban and rural districts by approaching the local school boards in the desired districts.
The data was collected in February and March of 2019.

The goal of research should be to gain an understanding of the relevant topic by applying the appropriate methods and techniques to the available data. By using the theory deductively, it was predicted that certain patterns would be visible based on the empirical evidence. A deductive approach involves using the theory as a foundation for establishing a hypothesis and performing the analysis based on the theoretical findings laid out in the literature review. This approach may generate valuable answers and is used in quantitative research to help predict events that may occur (Cresswell & Cresswell, 2018; de Vaus, 2014).

Through a quantitative approach, I sought to describe certain elements of reality which may not be immediately obvious when examining the data. A quantitative method can be regarded as a systematic approach to investigation of specific statistical scenarios. By processing numerical data in a computer program, it is possible to look for certain patterns that may give answers to specific research questions (Tolmie, Muijs, & McAteer, 2011). The study has a non-experimental and correlational design as the variables are not manipulated, leaving it difficult to draw conclusions about causality. Instead I used a cross-sectional approach which was descriptive and gave a snapshot of a group of children at one point in time. The data was examined through bivariate correlation and hierarchical regression analysis and sought to describe what the situation was at that specific point in time. A correlational design, allowed for description of the degree of relationship between the variables and inference about the relationships were drawn through multivariate regression analysis, although leaving room for uncertainty given the chosen design (Cresswell & Cresswell, 2018; Kleven, 2002b; Tabachnick & Fidell, 2007).

Descriptive statistics, bivariate correlation analysis and multivariate hierarchical regression analysis have been used for the analysis. Mahalanobis distances (MD) were examined to detect multivariate outliers and none of the variables had values greater than 15, where for samples of 100 and with fewer predictors is a cause for concern (Field, 2009). In the present study the highest value was 10. The histograms and skewness values were examined, and an outlier was detected in the text reading fluency variable. The outlier was verified in a scatterplot and a boxplot was used to identify the id number. The decision was made to winsorize the outlier to the next highest score plus one unit[4]. The remaining variables were deemed acceptable based on skewness values (Field, 2009).

The independent variables (IV) were chosen based on previous research and as the experimenter I have chosen to add them in a specific order (Field, 2009). By adding the components of text reading first I was able to predict the unique variance of receptive vocabulary (measured with PPVT) on oral text reading fluency (measured with a connected text). The analysis has been carried out in SPSS (IBM, SPSS Statistics 26).

3.1.1 Participants

The sample was drawn from grade 3 children. Normally developed readers in this age group in a relatively transparent orthography have established the component skills in reading and show differences in reading fluency.

As a researcher I want to draw conclusions about the whole population based on a smaller sample (Field, 2009). This study used a subset of a sample from a large project on orthographic learning which consisted of 73 children in grade 3. Children with learning difficulties were not filtered out of the sample and were able to conduct all the tasks. The project sample was considered representative of the population as it contained a proportionate

[4] Transformation using square root and logarithm was also attempted, but did not remove the outlier, hence the decision to replace the score.

number of students with minority background and learning difficulties. Cases with missing values were removed out of the analysis, leaving 70 children in the sample.

3.1.2 Procedure

The test assistants travelled to the schools and undertook the assessments in a quiet area during school hours. The test assistants all underwent the same level of training and each student was mainly tested by one assistant except for five students where testing was completed over the course of two sessions. The reading tasks used in the present study included one discrete word reading task, one serial word reading task and one serial digit naming task in addition to the one-page connected text and the receptive vocabulary measure.

3.2 Variables and assessment instruments

The variables were part of a larger battery of tests which took 45-60 minutes to administer. I will only be reporting on the assessments relevant to this study. The test battery consisted of normed tests and tests developed for research purposes.

Table 1

Overview of variables and assessment instruments

Variable	***Assessment instrument***
Oral text reading fluency	One page connected text
Receptive vocabulary	Peabody Picture Vocabulary Test (PPVT)
Individual word reading	Digital discrete word reading task
Serial word reading rate	Digital word list reading task
Serial processing rate	Digital digit naming task (RAN)

3.2.1 Assessment of oral text reading fluency

Oral text reading fluency was assessed using a Dutch connected one-page text consisting of 246 words in three paragraphs, where the whole text was read as fast and accurately as possible (Appendix 2; L. Bazen, personal communication, 2018). The text was a mix of narrative and expository and was somewhat unpredictable, meaning the child needed to pay attention whilst reading. Originally the text had been used as a silent reading task, hence the tasks throughout to check for understanding (i.e. “grab the yellow cube”). The child was instructed not to carry these out during the oral assessment. Omissions, insertions, mispronunciations, substitutions, and hesitations of more than 3 seconds were counted as errors. In the present study ORF was a time-based measure of accurate reading of a one-page connected text, where fluency was scaled as the number of words read correctly per second.

3.2.2 Assessment of receptive vocabulary

Receptive vocabulary was assessed using a Dutch version of the Peabody Picture Vocabulary Test (PPVT-NL; Schlichting, 2005). The task was administered according to the standardized procedure where the child is presented with a word orally and must select the appropriate picture out of four presented on a page. The test consists of 17 sets of 12 words. The raw score was the total number of correctly chosen pictures plus the autoscored pictures in the preceding sets depending on the child's age. To use homogenous measures, the PPVT raw score was used in the analysis as the other variables are not standardized. The reliability of PPVT-NL is considered good in a report by the Dutch Committee on Test and Testing (COTAN; Egebrink, Holly-Middelkamp, & Vermeulen, 2017; K. Vermeulen, personal communication, June 9, 2020).

3.2.3 Assessment of individual word reading

Individual word reading was assessed using a digital discrete word reading task where one word was presented on the screen at a time. A total of 36 high-frequency four letter words were administered. The child read the words aloud and the assessor clicked when the child was finished articulating it. Time was recorded offline using a graphical display of the recording. Instructions and practice tasks were provided before the test. The task was administered using the DMDX software (Forster & Forster, 2003). Reliability of the test was good (Cronbach's alpha = 0.98).

3.2.4 Assessment of serial word reading rate

Serial word reading rate was assessed using a digital word list reading task. A total of 36 high-frequency four letter words were displayed in four rows of nine words. The words were matched to the discrete task in terms of onset phoneme, length, consonant-vowel structure, and frequency. The child read the words aloud from top to bottom and total reading time of all the words was recorded. Instructions and practice tasks were provided before the test. The task was administered using the DMDX software (Forster & Forster, 2003). No reliability measures are available for this task as it was administered only ones.

3.2.5 Assessment of serial processing

Serial processing (RAN) was assessed using a rapid digit naming task. The child was presented with a matrix of 36 digits (nine repetitions of four digits displayed in 4 rows) and asked to name them as quickly and accurately as possible from top left to bottom right. The child was first presented with a practice task to ensure they understood the instructions given. The total score was the number of digits per second. The task was administered using the DMDX software (Forster & Forster, 2003). No reliability measures are available for this task as it was administered only once.

3.3 Validity and Reliability

In quantitative research numerical data is analysed and summarized systematically using statistical programs to detect patterns and relationships in the data to generalize from. Although a quantitative approach is believed to provide strong empirical evidence, it still has its weaknesses and limitations that need to be addressed. Thus, evaluating the study's validity and reliability is important (Cresswell & Cresswell, 2018; Tolmie et al., 2011).

Reliability of an assessment tool relates to whether it can be interpreted consistently across other settings. A measurement of a tool's internal consistency (Cronbach's alpha) is used for evaluation. It requires measurement of the same construct more than once. In the present study reliability measures are only available for PPVT-NL and discrete word reading and not for the other tests administered due to lack of test repetition.

Validity is related to evaluating the credibility of the inferences drawn from the quantitative analysis. The conclusions drawn from the results will never represent an absolute truth, but by evaluating threats to the validity of assessment tools and the statistical analysis chosen to answer the research question, it is possible to minimise the risk of conclusions being drawn on an incorrect basis (Field, 2009). In 1979, Cook and Campbell developed a framework for evaluating the validity of research which is also used in quantitative research and will form the basis for my evaluation of threats to validity. The types of validity are statistical conclusion validity, construct validity, internal validity and external validity (Cook & Campbell, 1979; Kleven, 2008; Lund, 2002).

3.3.1 Statistical conclusion validity

Statistical conclusion validity is especially relevant in this study due to its design. With statistical conclusion validity we evaluate whether a tendency is worthy of an interpretation or whether it is just a random incident. In quantitative research tests of effect size and significance are typically used to evaluate the strength of relationships. The significance level can be set at .05 or 5%, but the number is only arbitrary and stricter levels can be applied especially for larger samples. Generally, observed significance levels below the set value would be treated as if they were true. This involves a risk of Type I and Type II errors, meaning we reject a null-hypothesis (no relationship between variables) when in fact there is no relationship (Type I error) or we fail to reject the null-hypothesis when in fact there is a relationship (Type II error). In regression analysis statistical conclusion validity concerns whether the predictor and outcome variable covary and whether the conclusions drawn from the statistical analysis can actually be considered valid (Cook & Campbell, 1979; Kleven, 2008).

3.3.2 Construct validity

In quantitative research the constructs to be assessed, must be operationalised into measurable units. Construct validity relates to the extent to which the constructs that we are out to measure have been successfully operationalised. Do the measurement tools capture the essence of the constructs? Some concepts in educational research can be difficult to operationalise as they are more abstract. They may not be directly measurable, and we need to decide what is the most appropriate indicator of what we are trying to measure. To decide how best to measure a construct the researcher needs to be aware of what the best visible indicators are to capture the construct. Only then can construct validity be strengthened (Kleven, 2002a, 2008).

Construct validity would be affected by both random and systematic measurement errors. Random errors are difficult to alleviate but could be reduced by a large sample size and also by reliable measures. A consistent (reliable) measurement would give similar results if the same person were to be tested again. Systematic errors as a threat to construct validity can be improved by familiarity with assessment limitations and training of assessors to ensure consistent and correct use of instruments (Cook & Campbell, 1979; Kleven, 2008).

3.3.3 Internal validity

In quantitative research internal validity concerns being able to draw inferences from a covariation to a causal relationship, although in non-experimental designs establishing definite causality is not possible. To make any assumptions about causality, it would be necessary to rule out all other possible explanations. Internal validity in this study is weakened due to the lack of an experimental approach. An evaluation of this study's internal validity would therefore have to be based on assumptions made through theory to evaluate whether the chosen variables covary, the direction of the relationship and whether the variables have been entered correctly in the regression analysis. It would also be necessary to evaluate whether any variables not measured could cause a threat to internal validity. In terms of this study's research question the uncertainty about the direction of causal influence is a threat to internal validity as it may be that reading fluency also influences receptive vocabulary (Kleven, 2008; Lund, 2002).

3.3.4 External validity

External validity is concerned with being able to generalize from one sample to the population in terms of the children, situation, and context. In a non-experimental design, as with experiments in general the sample needs to be representative of the population and not drawn due to convenience. Heterogeneity in the sample will increase external validity if it matches heterogeneity in the population. In addition, low statistical conclusion validity is a threat to external validity as it affects the generalizability from the sample to the population (Cook & Campbell, 1979; Kleven, 2008).

Ecological validity is also an area to consider in educational research and is relevant to examine in relation to this study. Originally the terminology related to experiments conducted in a laboratory setting and considered whether the results could be deemed valid in terms of a natural setting. It is however important to consider ecological validity in terms of whether the actual assessment deviates from an activity undertaken in a natural setting, meaning if unfamiliar elements are added to the activity. It is suggested that a setting becomes ecologically valid when the researcher is aware of a subject's experience of the setting and when the experiment's intention corresponds with the environment in which the researcher wishes to generalize (Bronfenbrenner, 1979). In reference to this study it is important to consider whether both the setting and the assessment tools closely enough mimic an ordinary reading situation.

3.4 Ethical Considerations

The National committee for research ethics in the social sciences and the humanities (NESH) has developed ethical guidelines aiming to promote good scientific practice in research in Norway (De nasjonale forskningsetiske komiteene, 2016). As a researcher I adhere to the above guidelines and have considered these prior to and throughout the conduct of the present study.

Children are particularly vulnerable in research and steps must be taken to ensure respect for individuals and confidentiality when conducting the research and in analysing the results. Researchers must provide information on a project neutrally to avoid pressure and before starting a project. As the project deals with personal data informed consent has to be given in advance. Parents of children under the age of 15 have to give consent before letting the children take part in research (De nasjonale forskningsetiske komiteene, 2016). Parents in the Dutch project where the data in the present study originates were informed of the schools' participation in the project and had to give their consent on behalf of the children. The project was also approved by the Dutch ethics committee.

Children above the age of seven are entitled to be informed and have their opinion's heard. They should generally not be forced to participate in a project without giving their consent. Participants do have a right to withdraw and have all their data erased. This also applies to children where parents have given consent, but where they no longer wish to participate in a project. Researchers also need to respect participants privacy and storage of data must follow rules on data protection (De nasjonale forskningsetiske komiteene, 2016; Tolmie et al., 2011). Assessors in a study have a duty to perform the assessments without pressure and give sufficient feedback to the children throughout. Assessors have an obligation to see their own limitations and ensure they do not cause distress for the children in a test environment. All assessors in the project were trained in advance of performing any assessment to ensure they adhered to this practice throughout and did not upset the children (Tolmie et al., 2011).

When using data from another project steps must be taken to ensure the participants' anonymity. In the case of the present study all data has been anonymised and only id numbers have been provided. I have not had access to any personal information regarding the participants (De nasjonale forskningsetiske komiteene, 2016).

As a researcher I should be familiar with and comply with ethical norms. I adhere to good citation practice to avoid plagiarism and to promote transparency. The research must be presented truthfully. To my knowledge all sources have been credited appropriately. I have also taken steps not to misrepresent any data or results (De nasjonale forskningsetiske komiteene, 2016).

4. Results

Parametric tests have been applied in the data analysis using SPSS as the data is normally distributed, measured at the ratio level and the data is independent (Field, 2009).

4.1 Descriptive statistics

Table 2 gives a summary of information from the descriptive analysis. The table shows number of cases (*N*), median, mean, standard deviation (*SD*), skewness, kurtosis, and significance value for Kolmogorov-Smirnov test of normality.

Table 2

Descriptive statistics of variables

	N	*Median*	*Mean*	*SD*	*Skewness*	*Kurtosis*	*Kolmogorov-Smirnov, Sig.*
Text reading fluency[a]	70	2.06	2.05	0.53	.469	.350	.200*
PPVT[b]	70	113	112.86	9.53	-.356	-.066	.050
Age	70	106	106.17	4.71	-.068	-.712	.200*
Discrete WR[c]	70	0.96	0.94	.13	.337	-.177	.191
Serial WR[c]	70	1.76	1.73	.35	-.008	.091	.200*
Serial digit naming[c]	70	1.67	1.66	.32	.320	.181	.200*

Note. N = number of cases, SD= standard deviation, Sig = significance
[a] No. of correctly read words per second, [b] Raw score, [c] Items named per second
**this is a lower bound of true significance*

By looking at skewness and kurtosis it is possible to check the data's deviation from a normal distribution. Ideally values of skewness and kurtosis should be as close to zero as possible. Skewness values of zero indicate a perfectly symmetrical distribution. Positive values of skewness will indicate a pile-up on the left side of the distribution whilst a negative value will indicate a pile-up on the right-hand side. A positive kurtosis value will indicate a distribution with too many values in the tails whereas a negative value indicates a distribution with too few values in the tails. The relative concept of kurtosis being related to "peakedness" has been disputed and it is argued that kurtosis is largely related to the tails of the distribution and not the peak (Field, 2009; Westfall, 2014).

In addition to the histograms the Q-Q plots (Figure A1 to A6) show whether the observed data deviates from a normal distribution. If the data is normally distributed the dots fall along the line indicating the normal distribution (Field, 2009).

Cronbachs alpha (α) is a measure of the variable's reliability (internal consistency) and it ranges between 0 and 1 where a high number (above .7) makes the scale more reliable. By calculating α it is possible to check whether the tasks within a test correlate. In the present study α is only considered for PPVT-NL and discrete word reading as there was more than one task in the tests (de Vaus, 2014; Field, 2009).

4.2 Evaluating the variables

The Kolmogorov-Smirnov (K-S) test compares the study sample scores to that of a normal set of scores with the same mean and standard deviation to evaluate whether the distribution is close enough to a normal distribution. With $p \geq 0{,}05$ the test is non-significant, and the distribution of the sample is not significantly different from a normal distribution. Based on the K-S test results which indicate that all variables are non-significant ($p \geq 0.05$) it is assumed that the variables have an approximately normal distribution. (Table 2; Field, 2009). Below is a more detailed evaluation of the different variables.

4.2.1 Evaluating the variable text reading fluency[5]

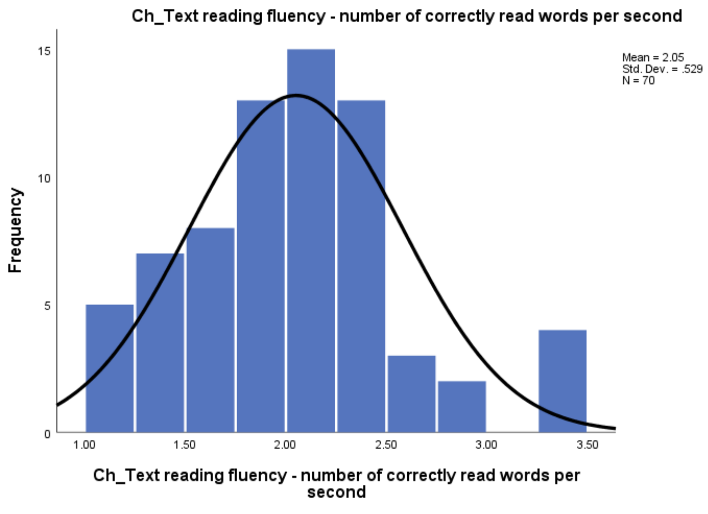

Figure 1. Histogram of text reading fluency scores

[5] The outlier in text reading fluency was changed to one unit above the second highest score. See section 3.1.

Text reading fluency has a positive skewness value (.469) which indicates a pile-up on the left side of the distribution, also visible from the histogram. A positive kurtosis value (.350) indicates a distribution with more values in the tails although the deviation from zero is small. The spread is considered close enough to a normal distribution after having performed the K-S test of normality indicating p = .200 (lower bound of true significance).

4.2.2 Evaluating the variable receptive vocabulary (PPVT)

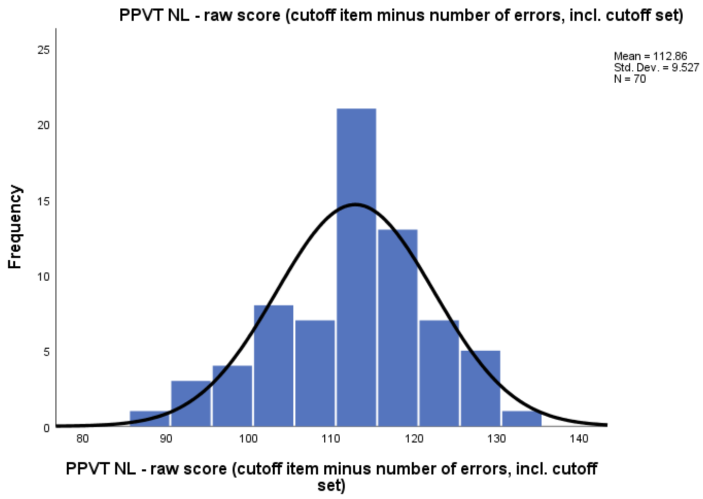

Figure 2. Histogram of PPVT scores

PPVT (receptive vocabulary) has a negative skewness value (-.356) indicating a slight pile-up to the right of the curve, also visible from the histogram. The kurtosis value (-.066) is close to zero and K-S test shows p = 0.05 which means the test is non-significant and the result close enough to a normal distribution.

4.2.3 Evaluating the variable age

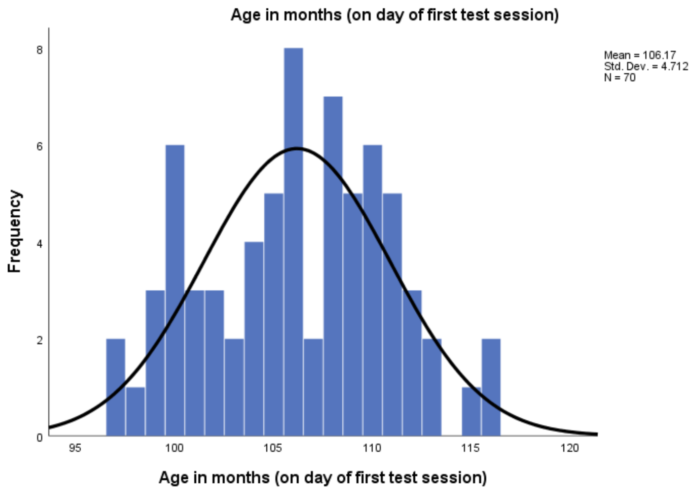

Figure 3. Histogram of age

Age has a slight negative skewness value (-.068) which is close to zero and a negative kurtosis value of (-.712) indicating fewer values in the tails. The K-S test shows p = .200 (lower bound of true significance) indicating that age is close enough to a normal distribution.

4.2.4 Evaluating then variable discrete word reading

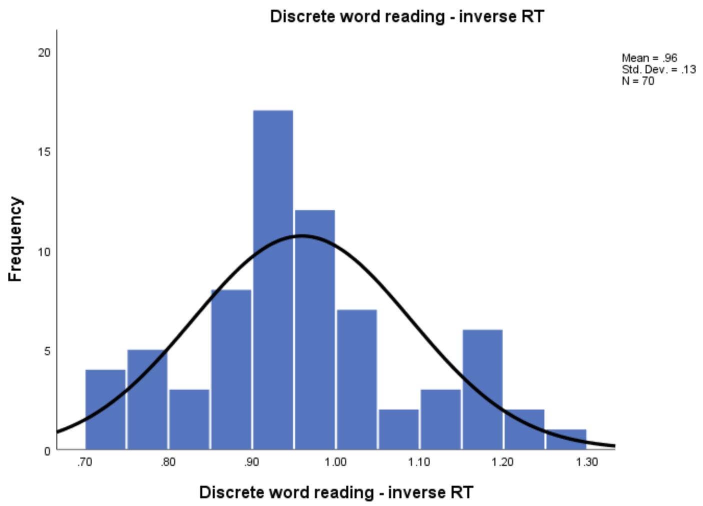

Figure 4. Histogram of discrete WR scores

Discrete word reading has a positive skewness value (.337) indicating a slight pile-up to the left of the curve and a negative kurtosis value (-.177) which is close to zero. K-S test of p = 0.191 shows that the sample does not deviate from a normal distribution.

4.2.5 Evaluating the variable serial word reading

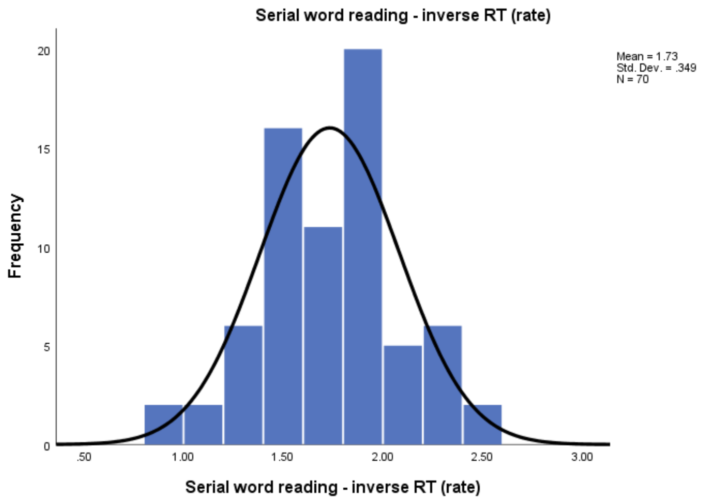

Figure 5. Histogram of WR scores

Serial word reading has a negative skewness value (-.008) and a positive kurtosis value (.091) both close to zero. K-S test of *p* = 0.200 (lower bound of true significance) shows that the sample does not deviate from a normal distribution.

4.2.6 Evaluating the variable serial digit naming (RAN)

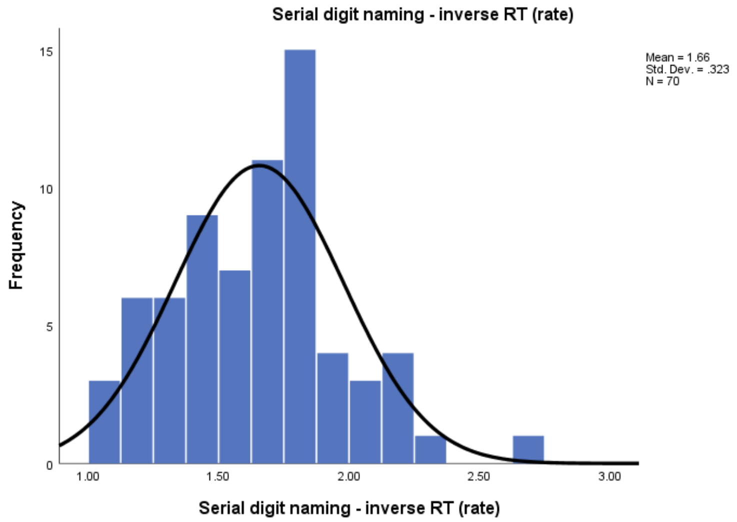

Figure 6. Histogram of serial digit naming

Serial digit naming (RAN) has a positive skewness value (.320) indicating a pile-up to the left side of the curve also visible in the histogram and a positive kurtosis value (.181), but relatively close to zero. K-S test indicate $p = 0.200$ (lower bound of significance) and the distribution is thereby not significantly different from a normal distribution.

4.3 Bivariate correlation analysis

A correlation coefficient describes the linear relationship between two variables. The relationship could be positive or negative. The correlation is expressed in the correlation coefficient: Pearson product-moment correlation coefficient *(r)*, which is used for parametric data and is considered a robust measure. The coefficient must be between +1 and -1. With a positive correlation, when one variable increases the other variable will also increase. If the coefficient is negative, then one variable will decrease as the other one increases. If the coefficient is zero, this indicates no relationship between the variables (Field, 2009). As a rule of thumb correlation coefficients of less than $\pm$ 0.1 are weak relationships, less than $\pm$ 0.3 are modest relationships, less than $\pm$ 0.5 are moderate relationships, less than $\pm$ 0.8 are strong relationships and equal to or greater than $\pm$ 0.8 are very strong relationships. Due to the complexity of educational processes and issues with measurement of variables, correlations in educational research are hardly ever strong or very strong (Tolmie et al., 2011). In this study the main interest is the correlation between receptive vocabulary and text reading fluency. By taking a preliminary glance at the scatterplot a relationship can be graphed between the data, which is a condition for conducting a regression analysis (Figure 7).

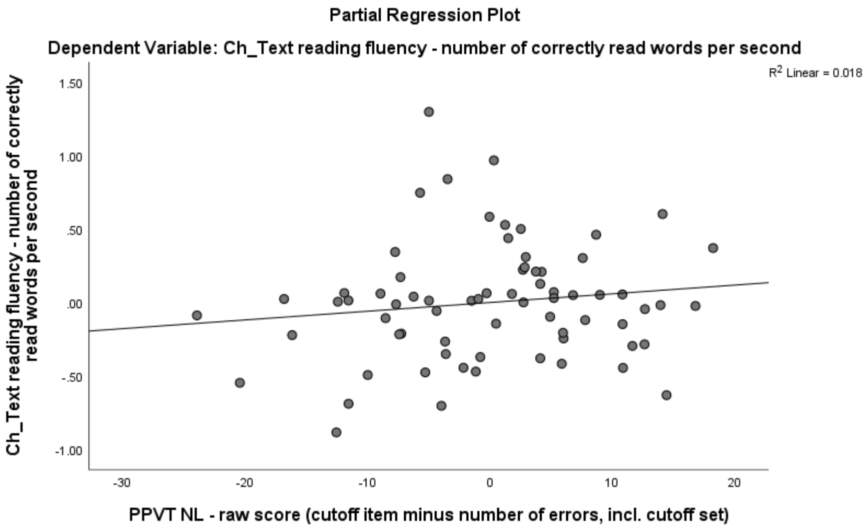

Figure 7. Scatterplot of relationship between receptive vocabulary and text reading fluency

The dots in Figure 7 represent individual cases. The scatterplot reflects a slightly positive relationship between receptive vocabulary and text reading fluency. The relationship does however seem to be weak (*r* = .12) and the dots have a slight spread (de Vaus, 2014). To further consider the correlations, Pearson's *r* and R-squared (R^2) is examined (Table 3). R^2 is the shared variance between the variables and indicates the proportion of the dependent variable (DV) that can be explained by the independent variable (IV) and can be expressed as a percentage.

When testing the assumption of a relationship between variables the null-hypothesis (H_0) is used. The H_0 in this case is that there is no correlation between the variables. A significance test in relation to correlation tells us the probability of getting a coefficient of the same size in a sample of 70 if there is no relationship between the variables in the population (the results could be due to sampling error). The significance test does not reveal anything about the strength of the relationship it just gives us a possibility to check how likely it is that the correlation could be present. The probability of the relationship occurring by chance is expressed in the *p*-value, a number between 0 and 1. The lower the significance level the less likely the H_0 is. A standard level of significance for smaller samples is 0.05, also used in this study. It means that there is there is a 5% chance that we might be wrongly rejecting H_0. For larger samples it is advisable to use a lower level of significance and 0.01 (99%) is often used (de Vaus, 2014; Field, 2009).

The correlation coefficients between the IVs: discrete word reading, serial word reading, serial digit naming and age are shown in Table 3. The correlation between them will not be addressed in any detail as they are used as control variables in this study. It is however interesting to examine how the IVs correlate with the DV, text reading fluency (in bold). Text reading fluency correlates significantly with discrete word reading and serial word reading at the .01 level which is unsurprising given that both measures tap processes involved in reading. There is a correlation at the .05 level with serial digit naming. The strongest correlation is with serial word reading (*r* = .647, equivalent to 42% (41.9%) shared variance). This is also unsurprising as serial word reading closely mimics text reading fluency. Discrete word reading has a correlation of .318, equivalent to 11% (10.6%) shared variance and serial digit naming a correlation of .235, equivalent to 6% (5.5%) shared variance. Age has a nonsignificant correlation with text reading fluency (*r* = 0.045, equivalent to a shared variance of 0.02%) which again is unsurprising given that the sample is small and from the same birth year.

There is a nonsignificant *(p* = .323) correlation between receptive vocabulary and text reading fluency. This means that we cannot reject the H_0 and rule out that the correlation happens by chance or is accounted for by serial word reading. Nonetheless the correlation coefficient *r* = .120, equivalent to 1.4% shared variance, is still a small finding.

Table 3

Correlation Analysis

	Text reading fluency	***Age***	***PPVT***	***Discrete WR***	***Serial WR***	***Serial digit naming***
Text reading fluency		**.077**	**.120**	**.318****	**.647****	**.235***
Age	0.5%		–.021	.248*	.244*	.160
PPVT	**1.4%**	.04%		–.148	.028	–.228
Discrete WR	**10.6%**	6.2%	2.2%		.416**	.252*
Serial WR	**41.9%**	5.9%	0.08%	17.3%		.443**
Serial digit naming	**5.5%**	2.6 %	5.2%	6.4%	19.6%	

Note. Pearson's r above the diagonal and R^2 in percent below
*** Correlation is significant at the .01 level. *Correlation is significant at the .05 level*

4.4 Hierarchical multivariate regression analysis

In conducting the multivariate regression analysis, a hierarchical approach has been used. In this approach the IVs have been entered in a specific order based on the above theoretical considerations. In model 1 only age was controlled for. In model 2 all the control variables thought to capture text reading (discrete word reading, serial digit naming and serial word reading) were added. In model 3 the predictor variable PPVT was added (Table 4). In regression analysis the IVs are often referred to as predictor variables and covariates (control variables) and the DV as the outcome variable (Field, 2009; Tabachnick & Fidell, 2007).

4.4.1 Objective

The aim of multivariate regression analysis is to estimate the effect of a predictor variable on an outcome variable and control for that effect using other variables. By controlling for the effect, we are simply able to look at the effect of only one predictor variable. The regression analysis does not establish causality, it merely reveals relationships amongst the variables, something which could also be caused by unmeasured variables Thus, it does not rule out the effect of other variables that are not included in the analysis (Field, 2009; Tabachnick & Fidell, 2007).

4.4.2 Assumptions

To draw conclusions in regression analysis, from a sample to the population, it is imperative to review a list of assumptions (Field, 2009).

The assumptions that the variables are *independent* (from different subjects), *quantitative and measured at least at the interval level* are met. The same applies to *non-zero variance*, meaning the variables do not have variances of zero. The partial regression plots in Figure A9 to A12 indicate a *linear* relationship between the residuals as the dots are randomly dispersed and do not curve. The residual scatterplot (Figure A8) show that the assumption of *homoscedasticity* is met as the dots are evenly dispersed around zero and do not seem to become wider (Field, 2009; Tabachnick & Fidell, 2007). *Multicollinearity* occurs when two variables are highly correlated which is a cause for concern when using regression analysis. If a second predictor variable is added to the model and accounts for the same variance as the first, it would be impossible to predict the unique variance of the second variable. As seen in the correlation matrix (Table 3) none of the variables are very highly correlated. (Pearson's r above 0.9). The variance inflation factors (VIF) are all well below 10 (they are around 1), the average VIF is around 1 (1.052) and the tolerance is above 0.2, all indicating that there is no cause for concern related to collinearity (Table 5; Field, 2009; Tabachnick & Fidell, 2007). The assumption of *independent errors* is checked with the Durbin-Watson test, which tests for correlations between the residuals. It is not applicable in the present study as the data are not time series so there is no seriality to be concerned about. The residuals should have a normal distribution. The histogram (Figure A7) indicates a slight pile-up to the left, but close enough to a *normal distribution of residuals*. Finally, the predictors should be uncorrelated with other external variables. Those variables are the ones that have not been included in the regression but could still influence the outcome variable. This assumption is difficult to meet in the present study as it is restricted by the variables available and the scope of the study. With this condition in mind the assumptions are met (Field, 2009; Tabachnick & Fidell, 2007).

4.4.3 Results

Table 4 and 5 give the summary of results and the coefficients from the multivariate regression analysis.

Table 4

Hierarchical multivariate regression analysis summary

Model	R^2	*Adjusted R^2*	*R^2 Change*	*F Change*	*df1*	*df2*	*p*
1	.006	-.009	.006	.406	1	68	.526
2	.433	.398	.427	16.339	3	65	.000
3	.444	.400	.010	1.187	1	64	.280

Model 1: age

Model 2: age, discrete WR, serial WR, RAN

Model 3: age, discrete WR, serial WR, RAN, PPVT

R^2 is the amount of variance that is explained by the model and R^2 change is the difference in the amount of variance from one model to the next (Field, 2009). Model 3 shows that receptive vocabulary explains a nonsignificant variance of 1.0% in text reading fluency (R^2 change = 0.01 or 1.0%). The analysis also shows that age, discrete word reading, serial word reading and serial digit naming account for about 40% of the total variance in text reading fluency (R^2 = 0.433 and adjusted R^2 = 0.398). Ideally R^2 and adjusted R^2 should be close as adjusted R^2 gives an idea of how well the sample generalizes. It tells us how much variance the predictors are accounting for if the model was derived from the population. The difference between R^2 and adjusted R^2 for model 3 is 0.044 or 4.4% meaning if the model was derived from the population it would account for about 4% less variance in text reading fluency, indicating a good cross-validity of this model (Field, 2009).

The results from ANOVA (Table A1) illustrate that the overall fit of model 3 is significant (F=10.2, $p < .001$) although the ANOVA results do not address individual predictions (Field, 2009).

A good model should have a large F-ratio, at least greater than one. The F-ratio is somewhat reduced between model 2 and 3 which indicates that adding receptive vocabulary has not necessarily improved the analysis of variance (Field, 2009).

As the predictors in model 2 are only added to remove the variance of text reading skills I am mainly interested in looking at model 3 (the final model) as it includes all the variables and gives an idea of the unique contribution of receptive vocabulary. The β-coefficient estimates how much the outcome variable will increase when the predictor variable increases by 1, holding all the other IVs constant. In other words, the β-coefficient is the relationship between each of the predictor variables and the outcome variable (text reading fluency). In this case the β-value for receptive vocabulary is close to 0.01 (0.006) indicating a positive relationship where for every one word the child understands (picture correct), text reading fluency (number of accurate words per second) increase with 0.01 words per second when the effects of the other IVs are held constant. The t-test helps us evaluate the predictors' contribution to the model. If p is significant then the predictor is making a significant contribution. The higher the value of t and the smaller the value of p the greater is the contribution of the predictor. In this case the contribution is non-significant (t = 1.09, p = .280). The only variable with a significant contribution is serial word reading (t = 5.62, $p < .001$) indicating that serial word reading accounts for both discrete word reading and serial naming.

To look at the individual contributions of the variables semi-partial correlations are evaluated (sr (part) and sr^2 in Table 5). The value of sr is squared and converted to a percentage to allow for interpretation of each variable's unique contribution on the total variance of the outcome variable. As model 3 only has one extra variable added (PPVT) sr^2 is equivalent to R^2 change, namely 1%. The control variables show that serial word reading unsurprisingly accounts for the largest portion of unique variance in text reading fluency (sr^2=27.5%), followed by discrete word reading (sr^2= 0.7%) and serial digit naming (sr^2= 0.1%) indicating that the latter two are captured by serial word reading in the model (Tabachnick & Fidell, 2007) .

Table 5

Coefficients

Model		β	*t*	*p*	*sr (part)*	*sr²*	*Tolerance*	*VIF*
1	(Constant)	1.132	.784	.435				
	Age in months	.009	.637	.526	.077		1.000	1.000
2	(Constant)	1.297	1.154	.253				
	Age in months	–.011	–.974	.334	–.091		.912	1.096
	Discrete WR	.330	.778	.439	.073		.799	1.251
	Serial WR	1.009	5.926	.000	.553		.692	1.444
	Serial digit naming	–.107	-.622	.536	–.058		.796	1.256
3	(Constant)	.552	.420	.676				
	Age in months	–.011	–.99	.326	–.092	.008	.912	1.096
	Discrete WR	.404	.94	.349	.088	.007	.779	1.284
	Serial WR	.973	5.62	.000	.524	.275	.667	1.499
	Serial digit naming	–.057	–.32	.750	–.030	.001	.743	1.346
	PPVT	.006	1.09	.280	.102	.01	.904	1.106

4.5 Summary

To sum up, the results of the analysis show that all variables have a near normal distribution allowing for Pearson's *r* to be used in the correlation analysis. The correlation analysis shows that text reading fluency (outcome) has a non-significant correlation with receptive vocabulary (predictor) ($r = .120$). As for the control variables receptive vocabulary has a non-significant correlation with serial digit naming ($r = –.228$) and a significant correlation with discrete word reading ($r = –.148$) and serial word reading ($r = .028$).

With relevance to the research question the positive correlation between receptive vocabulary and text reading fluency (Pearson's $r = .120$) indicates a weak relationship between the two, although we cannot reject the H_0 of no relationship existing. The shared variance between text reading fluency and receptive vocabulary is non-significant and close to 1.5% (1.4%).

There is a significant correlation between all three control variables capturing text reading, on .01 and .05 level. Age correlates significantly with serial word reading but has a non-significant correlation with discrete word reading and serial digit naming.

In the regression analysis there is a non-significant relationship between receptive vocabulary and text reading fluency. Receptive vocabulary can uniquely predict 1% of the variance in text reading fluency although the result is small and statistically indistinguishable from zero. There is no contribution from serial digit naming (0.7%) or discrete word reading (0.1%) once serial word reading has been added meaning serial word reading captures all processes involved in text reading in this particular study.

5. Discussion

The purpose of the present study was to examine whether there is a relationship between receptive vocabulary and oral text reading fluency through bivariate correlation and hierarchical regression analysis. The analysis used three components thought to be the underlying skills of text reading, namely individual word reading, serial word reading and serial processing of digits as control variables (in addition to age) to establish whether receptive vocabulary can uniquely explain variance in text reading fluency after controlling for the underlying components. As no studies are directly comparable in terms of the transparency of the language, the assessment tools or variables used in analysis, and little is still known with regards to the effect of various language skills on oral text reading fluency (Ouellette, 2006), the results are discussed in relation to tendencies found in other studies with reference to accurate word reading and comprehension as being connected to oral reading fluency.

5.1 Discussion: theoretical findings

Somewhat counterintuitively based on previous research of language skills and comprehension and the strong connection between ORF and comprehension (Fuchs et al., 2001; Jenkins et al., 2003; Ouellette, 2006; Tilstra et al., 2009), receptive vocabulary uniquely explains only 1% of the variance in oral text reading fluency. The finding is nonsignificant, meaning that statistically the results of the analysis are indistinguishable from zero.
There is a rather large unexplained amount of variance left after controlling for the independent variables related to text reading (57.3%), meaning other factors not measured in the present study are contributing to the variance in oral text reading fluency, which is to be expected.

As Dutch is a relatively transparent orthography, decoding skills become established earlier, thus language skills are likely to play a role earlier than in an opaque orthography like English (Lervåg et al., 2018). Consequently, the hypothesis that vocabulary should explain unique variance in oral text reading fluency in grade 3 does not seem unreasonable. However, as the results in the present study do not clearly demonstrate that, one could assume that this may be due to measurement error in terms of either the predictor or the outcome variables. In other words that one receptive vocabulary measure is not enough to capture vocabulary/language skills that may explain variance in fluency, or the text is too easy to actually impose any comprehension demands or to inhibit fluency. Instead it is a measure of accurate reading which should be well established by grade 3 in a relatively transparent orthography.

5.1.1 Language skills

Somewhat contradictive of the results in the present study Park and Uno (2015) established a connection between receptive vocabulary and fluency in their cross-sectional study (grade 1 to 4). Receptive vocabulary uniquely predicted context fluency in grade 2 and was a unique predictor of decoding and overall reading ability (word reading accuracy and fluency) across the grades, although the text reading fluency measure used was slightly different from the one in the present study. The text in Park and Uno was at a level where preschool children are able to read it accurately and errors were not recorded, suggesting that the passage used was a fairly easy rate measure whereas the text used in the present study is a fluency measure and appropriate in grade 3. This could possibly explain the discrepancies in the results. Although Hangul is considered a consistent (shallow) orthography, it is still a possibility that reading development is different in syllabic and alphabetic languages. Decoding skills are, as mentioned above, established earlier in shallow orthographies (Lervåg et al., 2018; Protopapas et al., 2012; Rakhlin et al., 2019), meaning other skills are relied upon for comprehension, consistent with the finding in Park and Uno where receptive vocabulary contributed to word reading accuracy in grade 1 to 4.

The direct effect of receptive vocabulary on word reading was also demonstrated in Ouellette (2006), although in this case receptive vocabulary explained unique variance in sight word reading rather than reading accuracy (decoding) as indicated in Park and Uno (2015). Ouellette also demonstrated the strong relation between semantic knowledge and comprehension where receptive vocabulary was subsumed by expressive vocabulary in regression analysis. The results do however indicate that a comprehensive battery of vocabulary tests may be a better indicator of overall language skills than just a pure receptive vocabulary measure as used in the present study. Although the Ouellette study does not specifically look at oral reading fluency of text, the connections between ORF, sight word reading and comprehension are strong as demonstrated in the literature review above (Fuchs et al., 2001; Jenkins et al., 2003; Tilstra et al., 2009), hence it may be feasible to assume that a more comprehensive test of vocabulary or semantic knowledge would explain more unique variance in oral text reading fluency than demonstrated in the present study. Clearly there is a challenge in separating the various vocabulary measures in that there could be noise from expressive vocabulary in receptive vocabulary measures and vice versa (Ouellette, 2006), yet, a study including more semantic measures and expressive vocabulary measures could potentially capture more of the variance in oral text reading fluency than was demonstrated in the present study.

Findings in Carretti et al. (2019) are more in line with the present study. Carretti et al. found vocabulary skills to have less of an impact in older readers. The study was conducted with readers in Italian, also considered to be a relatively transparent orthography. Although the children in Carretti et al. were older (grades 3-5, 6- 7 and 9-10) and may therefore not be directly comparable to the children in the present study, the results indicate that language skills do not have the influence on comprehension as expected in a relatively transparent language, at least not in relatively established readers. Carretti et al. did however only use comprehension measures (oral and silent tasks) to determine the level of reading ability and there were no direct measures of vocabulary skills as found in Ouellette (2006) which means that the two studies are not directly comparable. The Carretti et al. results do however reveal that for transparent languages the impact of vocabulary skills is possibly less than would be expected from research in for example English. In terms of the present study, if the sample consisted of relatively established readers, it is possible that the results indicate that they are not reliant on receptive vocabulary for text reading fluency.

However, findings in Carretti et al. (2019) are not in line with an earlier cross-sectional study in an opaque orthography (English) conducted by Jenkins et al. (2003) who found a high correlation (.83) between text understanding and text reading fluency in grade 4 children. This discrepancy may be due to the transparency of Italian as the words can be read accurately earlier than in English. Carretti et al. also omitted the influence of speed and oral reading which could explain the discrepancies in results.

Schwanenflugel et al. (2006) also found results contradicting Jenkins et al. (2003), in a cross-sectional study of English-speaking children in grades 1 to 3. The study found no mediating role of text reading fluency between word-reading and comprehension. It concluded that in the early stages of reading development children use their word-reading fluency skills for comprehension rather than text reading fluency. The children were however younger than in the Jenkins et al. study which may have caused contradicting results as the children could be too young to utilise the added benefit from context. If the text used for comprehension or the fluency text had been more demanding the results may have been different, as is also assumed in the present study.

The hypothesis in the present study is based on studies finding specific connections between vocabulary and sight word reading and vocabulary and comprehension, and also results from Jenkins et al. (2003) where context reading was evaluated specifically. The sample is similar in terms of age and the instruments used in the two studies are comparable. The Jenkins et al. study used specific fluency tasks on both context (passages) and context-free (word-list) measures. Jenkins et al. conducted the study with English speaking children in grade 4 who may be on par with grade 3 children in a transparent orthography in terms of reading development. Jenkins et al. found semantic knowledge to be the link between fluency and comprehension and to explain the difference in results between text and list reading. However, the finding in the present study does not support "the importance of context processes in reading fluency and reading comprehension" (Jenkins et al., 2003, p. 725). Jenkins et al. also supported the idea that language skills affecting comprehension also affect reading fluency of text. The discrepancies in the findings between the present study and Jenkins et al. may indicate that with a more complex text with higher comprehension demands the results in the present study could have been different.

5.1.2 Word reading, serial word reading, and RAN

The present study did not find individual (discrete) word reading to explain any unique variance in oral text reading fluency when added together with serial word reading (word lists) and serial digit naming (Table 4, model 2). Both individual word reading, and serial digit naming were subsumed by serial word reading in the regression analysis (serial word reading uniquely explains 27.5% of the variance). However, in a study with Russian children Rakhlin et al. (2019) found word reading to explain unique variance in fluency but the study did not include a serial word reading measure, thus it is difficult to directly compare.

As mentioned above Schwanenflugel et al. (2006) found no mediating role of text reading fluency for comprehension. The results suggest that serial processing of words in context is not important for comprehension in children of that age group, but that they rely on accurate decoding of individual words. However, it is interesting to note that in the present study serial processing subsumed decoding skills which did not explain any amount of significant unique variance in oral text reading fluency and could statistically have been left out of the regression. The discrepancy may be due to the difference in age between the subjects of the two studies and could also reflect the difference in transparency between the languages (Dutch and English).

It seems that serial word reading explains a significant amount of variance on its own without any substantial contribution from RAN or individual word reading when the text read is not too demanding. This is partially in line with Altani et al. (2019) who found RAN to be a significant predictor for text reading fluency, beyond isolated word reading speed in a transparent language (Greek) as early as in grade 3. Altani et al. did not control for serial processing of words but used it as an outcome variable. The results in the present study show that there is a moderate correlation between RAN and serial word reading (r = .443) supporting results in Altani et al. which showed that RAN is related to serial processing of text. Had serial word reading not been controlled for in the present study RAN would possibly have explained a larger amount of variance for grade 3 readers.

Protopapas et al. (2013) found a high association between vocabulary and decoding and comprehension in support of the lexical quality hypothesis where quality of word representations was important for skilled reading. The present study was not able to support this finding as serial word reading subsumed individual word reading and unique variance was not explained by the vocabulary measure in the present study.

Several studies have pointed to the importance of using fluency as a measure of overall reading competence especially once decoding is established (Fuchs et al., 2001; Jenkins et al., 2003; Language and Reading Research Consortium, 2015; Protopapas et al., 2013; Tilstra et al., 2009; Verhoeven et al., 2011). This is supported in the present study where serial word reading subsumes individual word reading indicating that once decoding is established as would be expected by grade 3 in a relatively transparent language (usually end of grade 1; Altani et al., 2019) other language aspects become important. The present study was however not able to support the idea of receptive vocabulary being the sole language skill to explain differences in oral text reading fluency.

5.1.3 Summary

In the present study there is a nonsignificant finding where receptive vocabulary uniquely predicts 1% of the variance in oral text reading fluency. 57.3% of the variance in oral text reading fluency is unexplained. The result is somewhat contradictive of findings in other related studies (Jenkins et al., 2003; Ouellette, 2006; Park & Uno, 2015), although the studies are not directly comparable. The present results supported findings in another transparent orthography although again comparison is a challenge as the age of the children was different (Carretti et al., 2019).

Serial word reading subsumes individual word reading and RAN in the regression analysis supporting research suggesting that once decoding is established other factors are important when reading in context (Language and Reading Research Consortium, 2015; Tilstra et al., 2009; Verhoeven et al., 2011).

5.2 Discussion: Validity and reliability

All research has limitations in terms of its validity and reliability and these limitations need to be addressed and evaluated. A researcher should be objective when evaluating the research conducted. Correlations between tools used to assess reading skills can help evaluating the validity of the various instruments used. Theoretical findings can help support the evaluation (Lund, 2002).

Reliability is a general threat to all four types of validity mentioned in the present study. It would affect construct validity in that a construct cannot be measured validly if the measurements are unreliable and measurement errors affect statistical conclusion validity, internal validity, and external validity (Lund, 2002). Reliability will mainly be touched upon under section 5.2.1. According to Cook and Campbell (1979) statistical conclusion validity affects the quality of the other three areas and is therefore evaluated first.

5.2.1 Statistical conclusion validity

Threats to statistical conclusion validity arise from either low statistical power or violation of statistical assumptions. Results from the statistical analysis should be critically evaluated in terms of covariations and statistical significance (Lund, 2002).

A power analysis involves estimating the size of the sample based on the desired magnitude given expected variances. In the present study a sample of 70 grade 3 children was already given and there was no option to include more children. It may be considered a weakness in the study that no power analysis was conducted, and the smaller sample size could have affected the results as mentioned above. In smaller samples statistical significance should not solely be relied upon as it can be more difficult to achieve (Cook & Campbell, 1979).

Assumptions for regression analysis are considered above (section 4.4.2). It is challenging to ensure that the predictors are uncorrelated with other external variables. There is a possibility that other external variables not included in the study could affect the outcome variable and have an influence on the result. For one, SES has not been used as a control variable and there are also other language measures that could have an influence on the result.

Low test reliability reduces power and weakens statistical conclusion validity (Lund, 2002). Only PPVT-NL and discrete word reading have information on reliability. Reliability of PPVT-NL are both considered to be good (section 3.2.3). The lack of reliability estimates in some of the instruments are a weakness in the present study, although the types of instruments used to test serial word reading, and sequential processing (RAN) are widely used in research, strengthening the statistical conclusion validity (Altani et al., 2019; Carretti et al., 2019; Jenkins et al., 2003; Tilstra et al., 2009; Verhoeven et al., 2011). Measurement errors are random errors and belong under statistical conclusion validity in Cook and Cambell's (1979) system. But measurement errors would also affect construct validity and internal and external validity.

The findings in the present study are not statistically significant thereby making it challenging to draw conclusions based on the results which naturally weakens statistical conclusion validity. However, the analysis has been carried out dutifully and results have been critically evaluated giving statistical conclusion validity more strength.

5.2.2 Construct validity

To evaluate whether the measurement of the construct is valid and actually measures what is intended it is necessary to look back at theory and how the constructs are defined and generally measured in research. Correlations between measurements intended to measure reading skills can help verify construct validity (Kleven, 2002a).

Oral text reading fluency is generally measured when reading a text (or word lists) aloud for a period of 45s to 60s where errors are recorded (Carretti et al., 2019; Jenkins et al., 2003;

Schwanenflugel et al., 2006; Verhoeven et al., 2011). This is thought to capture overall reading ability (Fuchs et al., 2001). A text is often either narrative or expository. The score in the present study is a fluency score of number of accurate words read where total reading time for the one-page text is recorded and inverted to words correct per second. The text design (tasks throughout as it was originally a silent reading task) and the mix of narrative/expository need to be questioned as it may have created some level of noise in terms of the child's usual reading habit (see section 5.2.4). Using several texts and the mean number of words per second could have been a better alternative and is often done in normed fluency tests (Arnesen et al., 2017). The text does however have a significant and strong correlation with the word list task (r = .647) which speaks in favour of the validity of both tasks as measuring similar constructs.

The construct validity of PPVT-R, the American version and the one PPVT-NL is based on is considered good, thus supporting the notion that PPVT-NL is a valid measure of receptive vocabulary (Miller & Lee, 1993). The test does not require you to use language and is thought to be a good measure of words you perceive. The test is standardised, and reliability is considered good (see section 3.2.2).

Discrete word reading is generally used as a measure of isolated word reading (Altani et al., 2019) a prerequisite for developing skilled reading. In the present study discrete word reading has a significant moderate correlation with serial word reading and text reading fluency (r = .416 and .318 respectively) at the .01 level displaying the relationship between individual word reading and measures of word reading in sequence. Theoretical findings support the strong relation between individual word reading and serial word reading, thus the moderate correlation strengthens the validity of the test.

Serial word reading assesses the ability to process words in sequence and should be a good indicator of overall reading ability (Fuchs et al., 2001). In the present study serial word reading is a rate measure and accuracy was not recorded as the words were at a level where they are read accurately in grade 3. The correlation between serial word reading and text reading fluency was, as mentioned above, strong, indicating good validity of the test as a measure of overall reading competence. Serial digit naming is used as a rate measure of sequential processing and has been found to have a moderate correlation with reading ability (r = .43; Araújo et al., 2015). The correlation between serial digit naming and serial word processing (word lists) has been found to be strong in previous studies (Altani et al., 2019; Protopapas et al., 2013). In the present study serial digit naming has a significant moderate correlation with serial word reading (*r* = 443) at the .01 level and a modest correlation with text reading fluency (*r* = .235) at the .05 level.

Construct validity in the present study is considered strong, although the design of the fluency task may weaken validity of that particular test.

5.2.3 Internal validity

As this study has a non-experimental design internal validity is naturally weakened. The random selection of participants does however strengthen internal validity despite the lack of an experimental design (Kleven, 2002a).

In the present study bivariate correlation and multivariate regression analysis have been used to analyse the data. The correlation analysis addresses the covariance between variables but does not reveal causality. Thus, if the objective of the study is to address causal effects, either a different design should be chosen, or theoretical findings can be applied in an attempt to define the causal relationship by eliminating other possible explanations. It would however not be possible to be certain about these conclusions (Kleven, 2002b). However, as the aim of the present study was to look at the degree of variance explained by receptive vocabulary in oral text reading fluency it was not necessary to draw conclusions about causal effects, meaning the chosen method for analysis was sufficient, although it limits the conclusions available in terms of educational implications.

Using multivariate regression analysis to answer the research question can also strengthen internal validity as it gives an opportunity to address more specifically the contribution of the various independent variables. As mentioned above there is however a risk that important variables are left out of the analysis or that measurement error could have affected the result in some variables (Kleven, 2002b).

There could also be a threat of selection bias as there is a chance that some children have been left out of the research if they did not get permission by parents. This can however not be verified. Experimenter bias could be a threat too as the assessor could have been behaving in different ways with the children they were assessing. This is difficult to verify, but the assessors did receive training and the children mainly had one assessor each who carried out the whole test battery.

Internal validity is gauged as low in the present study mainly due to its design. There is no opportunity to determine causality nor look at the data longitudinally. The design does however give a snapshot of a specific point in time in a group of grade 3 children, thus strengthening validity for the intended purpose of the study.

5.2.4 External validity

External validity relates to whether the results would apply also to other people, time, and situations. A study with good external validity makes it possible to generalize to the population. The sample should be random to ensure it is representative of the population and to allow for generalizability (Lund, 2002). The sample in the present study has been randomly picked from different school districts and is meant to represent a normal variation in terms of reading ability and SES. Children with minority background and reading difficulties have not been filtered out. As mentioned above a weakness in the present study could be due to the sample size. A sample of 70 may not be enough to detect an effect and could be the reason for the results of the nonsignificant weak effect. The sample size could also cause an issue with generalization.

Ecological validity has also been evaluated due to the text design. The text had a format which can be described as a mix of narrative and expository. As mentioned, it had small tasks throughout as it was originally used to assess silent reading. This may have confused the children even though they were instructed prior to assessment not to carry out the various tasks. Yet the children’s usual reading behaviour may have been affected and the results not necessarily portraying their typical oral reading behaviour. As the text seemed to deviate from a more normal narrative/expository text (real-world text) it may not be possible to generalize from the findings in the study to the population. However, if the text is more difficult to read due to the tasks holding the reader up, they should have had low scores. The text reading fluency histogram (Figure 1) does indicate a pile-up on the left side of the distribution in support of this, but it does not seem to be reflected in the regression results. With a more difficult text one would assume that receptive vocabulary would explain a significant amount of unique variance in text reading fluency as the readers would rely on other abilities than decoding, however, this is not reflected in the results. Again, external (ecological) validity could have been strengthened if more texts were read in a natural setting and the mean score was used.

External validity is gauged as low in terms of the dependent variable possibly being affected by the design of the text. Otherwise external validity is sufficiently strong as the study can be replicated and the tasks have been carried in a natural environment at the respective schools.

5.2.5 Summary

The present study's internal validity is somewhat weak due to its design and does not allow for causal conclusions. Construct validity is considered relatively strong in this study, although only one measure has been used for each construct. It could further strengthen construct validity to use more than one fluency measure. Based on previous research, elements of reading in sequence seem well captured in the three variables testing this aspect. Statistical conclusion validity is also a strength in this study in terms of how the analysis has been carried out, although a larger sample could have made the conclusions more valid and possible also given different results. The sample was randomly chosen strengthening external validity although the text design has a weakness in terms of it originally being a silent reading task which makes it more difficult to generalize from (Lund, 2002).

The present findings were somewhat unexpected considering the theoretical findings above and by replicating the study with a larger sample, a more demanding text (or more texts) and/or more vocabulary measures results could possibly be different.

5.3 Limitations and implications for future research and education

5.3.1 Limitations

The present study addresses a gap in reading research by looking at the relationship between an area of vocabulary and oral reading fluency in context. Due to the scarcity of studies looking directly at this relationship it is somewhat challenging to compare results. There is however a discrepancy between the present study and other studies looking at language skills and fluency/comprehension. The small and nonsignificant finding in the present study could be due to the chosen variables or a relatively small sample. It has been highlighted above that using more extensive language measures or a different text could lead to different results. The importance of applying a larger battery to establish a more complete picture of children's overall language skills have been recognized in several studies (Braze et al., 2016; Lervåg et al., 2018; Protopapas et al., 2012). The rather large unexplained amount of variance in the fluency task suggests that the inclusion of more measure may have altered the results. Naturally, the transparency of the language complicates generalization to more opaque languages like English. Hence, more research is also needed in other transparent languages and in English to identify the role of vocabulary skills in oral text reading fluency.

5.3.2 Educational implications

Research supports the idea that oral reading fluency should be assessed as part of overall reading assessment in schools. Skilled reading can be assessed using oral reading fluency measures as an indicator of overall reading competence (Fuchs et al., 2001; National Reading Panel, 2000). As vocabulary breadth and depth also have been found to have implications for reading, enhancing, and assessing vocabulary should also be incorporated in education, in particular expansion of the semantic system (Ouellette, 2006).

The current study did not establish a significant finding in terms of receptive vocabulary predicting oral text reading fluency. However, the small finding is not to be completely neglected and supports the notion of a relationship. The weak relationship may as highlighted above, be due to the sample size or assessment instruments applied.

The study design has limitations and makes it difficult to address the implications for educational practices as causality is not established. Yet based on previous research one would assume that by being exposed to more words, reading improves and by increasing vocabulary, reading fluency in context will also improve (Jenkins et al., 2003; Ouellette, 2006).

5.3.3 Implications for future research

The present study only gave a snapshot of grade 3 children in a relatively transparent language at a specific point in time. As reading develops over the course of primary school years, it could be interesting to look at similar skills longitudinally to determine the specific impact of early vocabulary skills on oral reading fluency later and vice versa. In addition, an experimental design can help establish causality which would have an impact on educational practices.

Future studies could address the sample size and also the age being investigated. It has also been suggested an increase in the language constructs being assessed and inclusion of a more demanding text or more texts for the fluency measure. It could also be of interest to include specific comprehension measures to verify Fuchs et al. (2001) idea of fluency being an indicator of reading ability, also in a transparent language.

In terms of the chosen method, future studies could also run the analysis the opposite direction to determine the influence of oral text reading fluency on vocabulary/language skills as studies have indicated the reciprocal relationship between fluency and vocabulary (Jenkins et al., 2003; Verhoeven et al., 2011)

A rather large proportion of the variance in oral text reading fluency remained unaccounted for. To address this, other studies should include more language measures, like grammar, expressive vocabulary, semantic knowledge, and possibly also other cognitive measures like verbal working memory.

6. Conclusion

The aim of the present study was to examine the role of receptive vocabulary knowledge in oral text reading fluency in grade 3 children. The results of the regression analysis show that a child's range of receptive vocabulary knowledge can uniquely explain only 1% of the variance in oral text reading fluency over and above serial word reading rate and the finding is not statistically significant. However, future studies may be able to detect a larger connection by including more measures. Serial word reading measures may be sufficient to explain word reading skills at this age in reading development and the rather large amount of variance unaccounted for suggests that there are other skills that need to be included in future research.

Other studies tend to show a connection between the range of language skills and reading competence indicating that being able to uniquely explain a relationship as specified in my research question and hypothesis should be statistically possible by changing the outcome variable or the predictors.

List of references

Altani, A., Protopapas, A., Katopodi, K., & Georgiou, G. K. (2019). From individual word recognition to word list and text reading fluency. *Journal of Educational Psychology, 112*, 22–39. https://doi.org/10.1037/edu0000359

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology*, *107*, 868–883. https://doi.org/10.1037/edu0000006

Arnesen, A., Braeken, J., Baker, S., Meek-Hansen, W., Ogden, T., & Melby-Lervåg, M. (2017). Growth in oral reading fluency in a semitransparent orthography: Concurrent and predictive relations with reading proficiency in Norwegian, grades 2–5. *Reading Research Quarterly*, *52*, 177–201. https://doi.org/10.1002/rrq.159

Braze, D., Katz, L., Magnuson, J. S., Mencl, W. E., Tabor, W., Van Dyke, J. A., … Shankweiler, D. P. (2016). Vocabulary does not complicate the simple view of reading. *Reading and Writing*, *29*, 435–451. https://doi.org/10.1007/s11145-015-9608-6

Bronfenbrenner, U. (1979). *The ecology of human development. Experiments by nature and design*. Massachusetts: Harvard University Press.

Burger, A., & Chong, I. (2011). Receptive vocabulary. In S. Goldstein & J. A. Naglieri (Eds.), *Encyclopedia of child behavior and development*. Retrieved from https://link.springer.com/referenceworkentry/10.1007/978-0-387-79061-9_2359

Byrnes, J. P., & Wasik, B. A. (2019). *Language and literacy development. What educators need to know.* Retrieved from https://books.google.no/books?hl=en&lr=&id=4GehDwAAQBAJ&oi=fnd&pg=PP1&dq=Language+and+literacy+development:+What+educators+need+to+know&ots=-DZpgwFkeO&sig=OjD0LY_5Wgno5OftjUnxGPgx9ss&redir_esc=y#v=onepage&q=

Language%20and%20literacy%20development%3A%20What%20educators%20need%20to%20know&f=false

Carretti, B., Toffalini, E., Saponaro, C., Viola, F., & Cornoldi, C. (2019). Text reading speed in a language with a shallow orthography benefits less from comprehension as reading ability matures. *British Journal of Educational Psychology*, *90*, 91–104. https://doi.org/10.1111/bjep.12307

Catts, H. W. (2018). The simple view of reading: Advancements and false impressions. *Remedial and Special Education, 39*, 317–323. https://doi.org/10.1177/0741932518767563

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation. Design & analysis issues for field settings*. Boston: Houghton Mifflin Company.

Cresswell, J. W., & Cresswell, J. D. (2018). *Research design. Qualitative, quantitative & mixed method approaches*. California: Sage Publications.

De nasjonale forskningsetiske komiteene. (2016). *Forskningsetiske retningslinjer for samfunnsvitenskap, humaniora, juss og teologi [Guidelines for research ethics in social sciences, humanities, law and technology]* (No. 4). Retrieved from https://www.etikkom.no/forskningsetiske-retningslinjer/Samfunnsvitenskap-jus-og-humaniora/

de Vaus, D. (2014). *Surveys in social research*. Abingdon: Routledge.

Egebrink, I. J. L., Holly-Middelkamp, F. R., & Vermeulen, C. S. M. (2017). *COTAN beoordeling 2008, vragenlijst voor gedragsproblemen bij kinderen [COTAN review 2008, disruptive behaviour disorder rating scale].* Retrieved from COTAN NIP website: https://www.cotandocumentatie.nl/beoordelingen/b/13623/peabody-picture-vocabulary-test-iii-nl/

Ehri, L. (1998). Grapheme–phoneme knowledge is essential for learning to read words in English—word recognition in beginning literacy. In J. Metsala L. & L. Ehri (Eds.), *Word recognition in beginning literacy*. Retrieved from https://learning.oreilly.com/library/view/word-recognition-in/9780805828986/xhtml/07_chapter01.xhtml

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

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

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116–124. https://doi.org/10.3758/BF03195503

Fuchs, L. S., Fuchs, D., Hosp, M., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*, 239–256. https://doi.org/10.1207/S1532799XSSR0503_3

García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, *84*, 74–111. https://doi.org/10.3102/0034654313499616

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*, 6–10. https://doi.org/10.1177/074193258600700104

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*, 127–160. https://doi.org/10.1007/BF00401799

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95*, 719–729. https://doi.org/10.1037/0022-0663.95.4.719

Kleven, T. A. (2002a). Begrepsoperasjonalisering. In T. Lund (Ed.), *Innføring i forskningsmetodologi [Introduction to research methodology]* (pp. 141–184). Bergen: Fagbokforlaget.

Kleven, T. A. (2002b). Ikke-eksperimentelle design. In T. Lund (Ed.), *Innføring i forskningsmetodologi [Introduction to research methodology]* (pp. 265–286). Bergen: Fagbokforlaget.

Kleven, T. A. (2008). Validity and validation in qualitative and quantitative research. *Nordic Studies in Education*, *28*, 219–233. Retrieved from https://www-idunn-no.ezproxy.uio.no/file/pdf/33192817/validity_and_validation_in_qualitative_and_quantitative_research.pdf

Kuhn, M. R., Schwanenflugel, P. J., Meisinger, E. B., Levy, B. A., & Rasinski, T. V. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly*, *45*, 230–251. https://doi.org/dx.doi.org/10.1598/RRQ.45.2.4

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

Language and Reading Research Consortium. (2015). Learning to read: Should we keep things simple? *Reading Research Quarterly*, *50*, 151–169. https://doi.org/10.1002/rrq.99

Language and Reading Research Consortium. (2017). Oral language and listening comprehension: Same or different constructs? *Journal of Speech, Language & Hearing Research*, *60*, 1273–1284. https://doi.org/10.1044/2017_JSLHR-L-16-0039

Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development*, *89*, 1821–1838. https://doi.org/10.1111/cdev.12861

Logan, G. D. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. *Reading & Writing Quarterly*, *13*, 123–146. https://doi.org/10.1080/1057356970130203

Lund, T. (2002). Metodologiske prinsipper og referanserammer. In T. Lund (Ed.), *Innføring i forskningsmetodologi [Introduction to research methodology]* (pp. 265–286). Bergen: Fagbokforlaget.

Miller, Linda T., & Lee, C. J. (1994). Construct validation of the Peabody Picture Vocabulary Test—Revised: A structural equation model of the acquisition order of words. *Psychological Assessment*. https://doi.org/10.1037/1040-3590.5.4.438

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: reports of the subgroups* (No. NIH Pub. 00-4769). Retrieved from https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, *98*, 554–566. https://doi.org/10.1037/0022-0663.98.3.554

Park, H.-R., & Uno, A. (2015). Cognitive abilities underlying reading accuracy, fluency and spelling acquisition in Korean Hangul learners from grades 1 to 4: A cross-sectional study. *Dyslexia*, *25*, 235–253. https://doi.org/10.1002/dys.1500

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

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

Pikulski, J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher*, *58*, 510–519. https://doi.org/10.1598/RT.58.6.2

Protopapas, A., Katopodi, K., Altani, A., & Georgiou, G. K. (2018). Word reading fluency as a serial naming task. *Scientific Studies of Reading*, *22*, 248–263. https://doi.org/10.1080/10888438.2018.1430804

Protopapas, A., Mouzaki, A., Sideridis, G., Kotsolakou, A., & Simos, P. (2013). The role of vocabulary in the context of the simple view of reading. *Reading and Writing Quarterly*, *29*, 168–202. https://doi.org/10.1080/10573569.2013.758569

Protopapas, A., Simos, P. G., Sideridis, G. D., & Mouzaki, A. (2012). The components of the simple view of reading: A confirmatory factor analysis. *Reading Psychology*, *33*, 217–240. https://doi.org/10.1080/02702711.2010.507626

Rakhlin, N. V., Mourgues, C., Cardoso-Martins, C., Kornev, A. N., & Grigorenko, E. L. (2019). Orthographic processing is a key predictor of reading fluency in good and poor readers in a transparent orthography. *Contemporary Educational Psychology*, *56*, 250–261. https://doi.org/10.1016/j.cedpsych.2018.12.002

Schlichting, L. (2005). *Peabody picture vocabulary test-III-NL*. Amsterdam: Hartcourt Assessments B.V.

Schwanenflugel, P. J., Meisinger, E. B., Wisenbaker, J. M., Kuhn, M. R., Strauss, G. P., & Morris, R. D. (2006). Becoming a fluent and automatic reader in the early elementary school years. *Reading Research Quarterly*, *41*, 496–522. https://doi.org/10.1598/RRQ.41.4.4

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*, 143–174. https://doi.org/10.1348/000712603321661859

Stanovich, K. E. (2000). *Progress in understanding reading. Scientific foundations and new frontiers*. New York: The Guildford Press.

Sveen, A. (2011). Semantikk. In A. Sveen (Ed.), *Språk. En grunnbok [Foundations of Language]* (pp. 64–94). Oslo: Universitetsforlaget.

Tabachnick, B., G., & Fidell, L., S. (2007). *Using multivariate statistics*. Boston: Pearson International Education.

Tilstra, J., McMaster, K., Broek, P. V. den, Kendeou, P., & Rapp, D. (2009). Simple but complex: Components of the simple view of reading across grade levels. *Journal of Research in Reading*, *32*, 383–401. https://doi.org/10.1111/j.1467-9817.2009.01401.x

Tolmie, A., Muijs, D., & McAteer, E. (2011). *Quantitative methods in educational and social research using SPSS*. Berkshire: Open University Press.

Utdanningsdirektoratet. (2017). *Rammeverk for grunnleggende ferdigheter [Framework for core skills]*. Retrieved from https://www.udir.no/laring-og-trivsel/rammeverk/rammeverk-for-grunnleggende-ferdigheter/

Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading*, *15*, 8–25. https://doi.org/10.1080/10888438.2011.536125

Westfall, P. H. (2014). Kurtosis as peakedness, 1905–2014. R.I.P. *The American Statistician*, *68*, 191–195. https://doi.org/10.1080/00031305.2014.917055

Appendix 1, Table and figures

Normal Q-Q plots

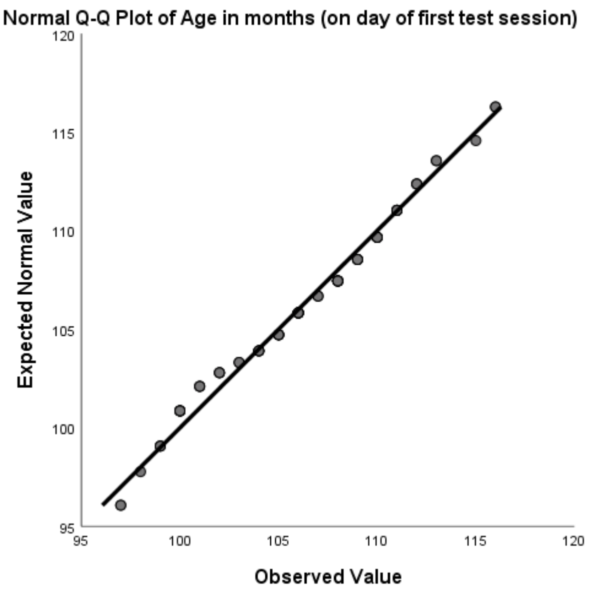

Figure A1. Normal Q-Q plot of age

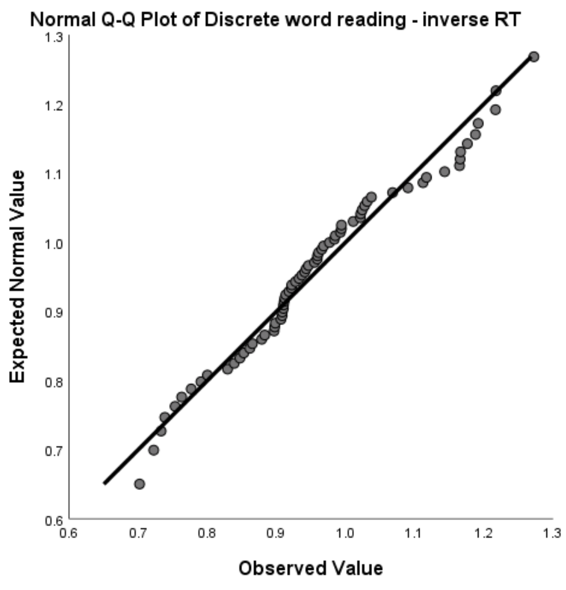

Figure A2. Normal Q-Q plot of discrete WR

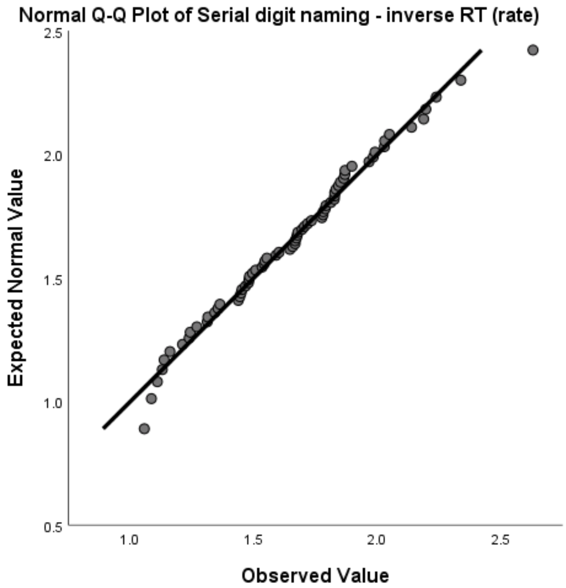

Figure A3. Normal Q-Q plot of serial digit naming

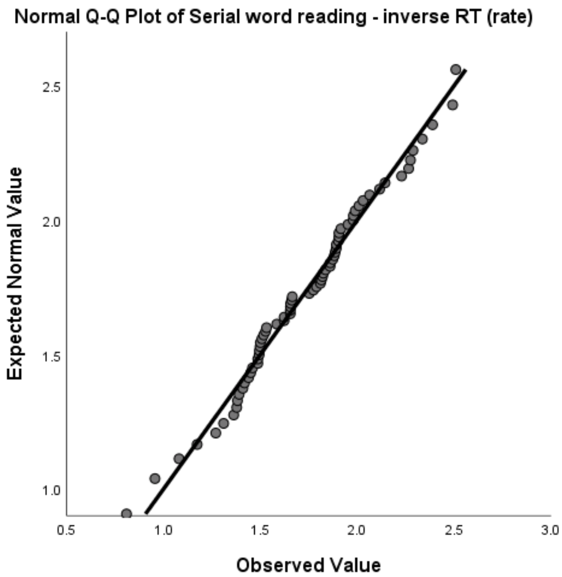

Figure A4. Normal Q-Q plot of serial WR

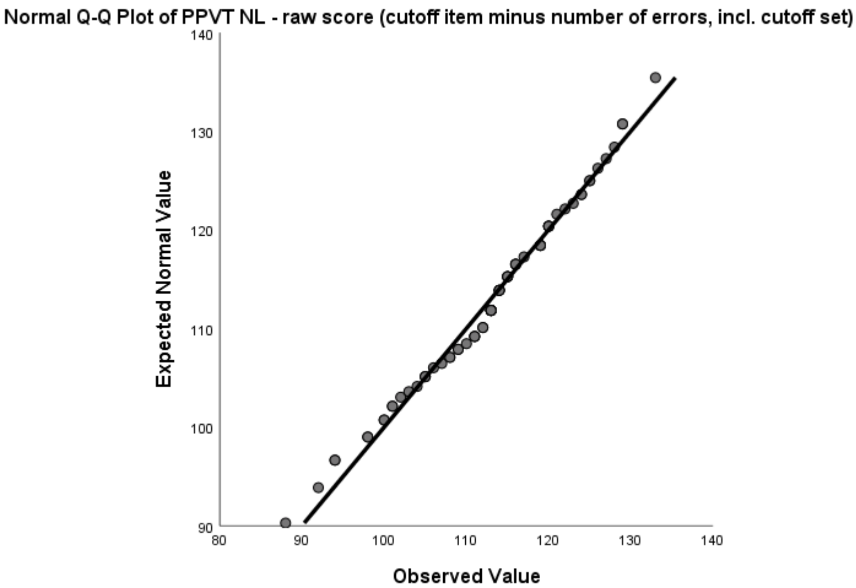

Figure A5. Normal Q-Q plot of PPVT

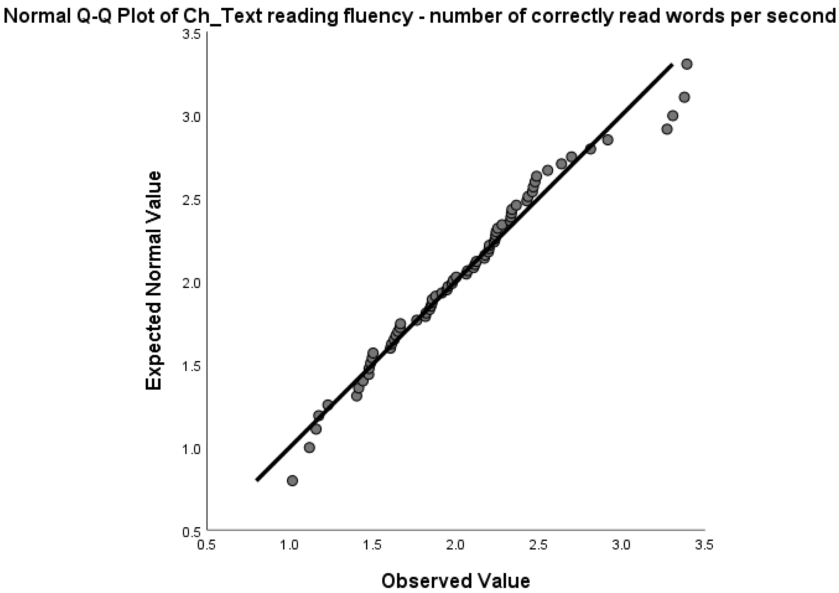

Figure A6. Normal Q-Q plot of text reading fluency

Histogram of residuals in Text Reading Fluency

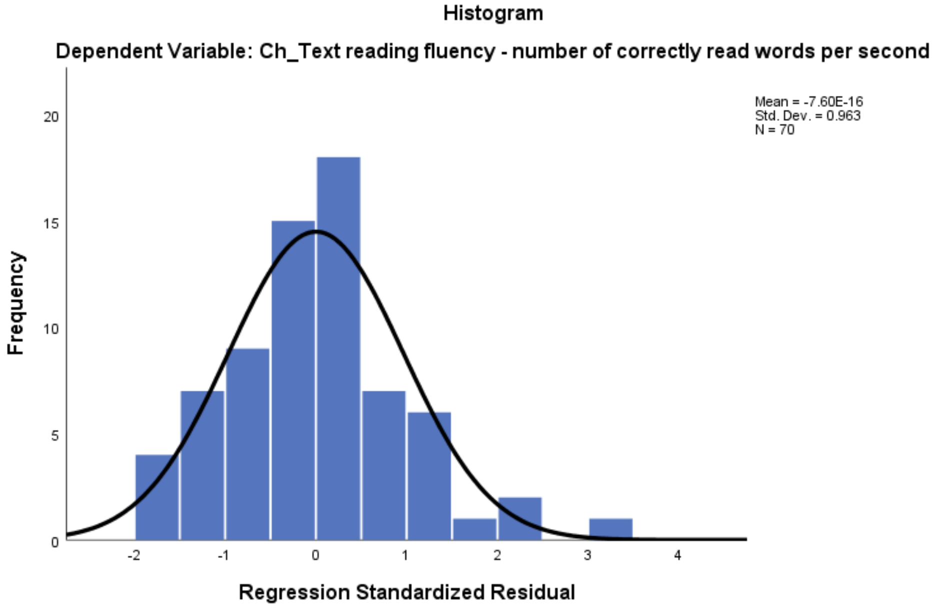

Figure A7. Histogram of residuals, text reading fluency

Scatterplot of residuals

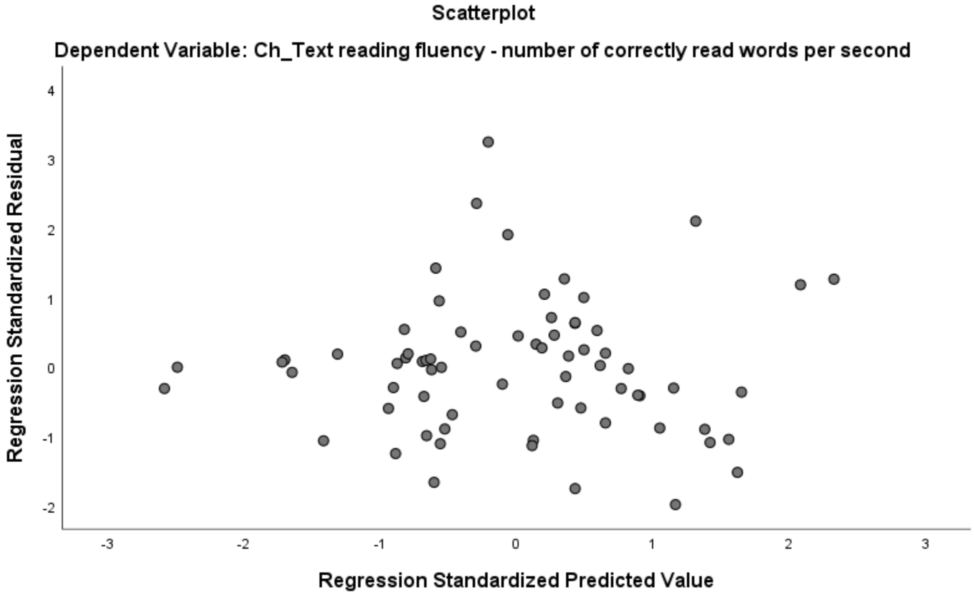

Figure A8. Scatterplot of residuals, text reading fluency

Partial Regressions Plots

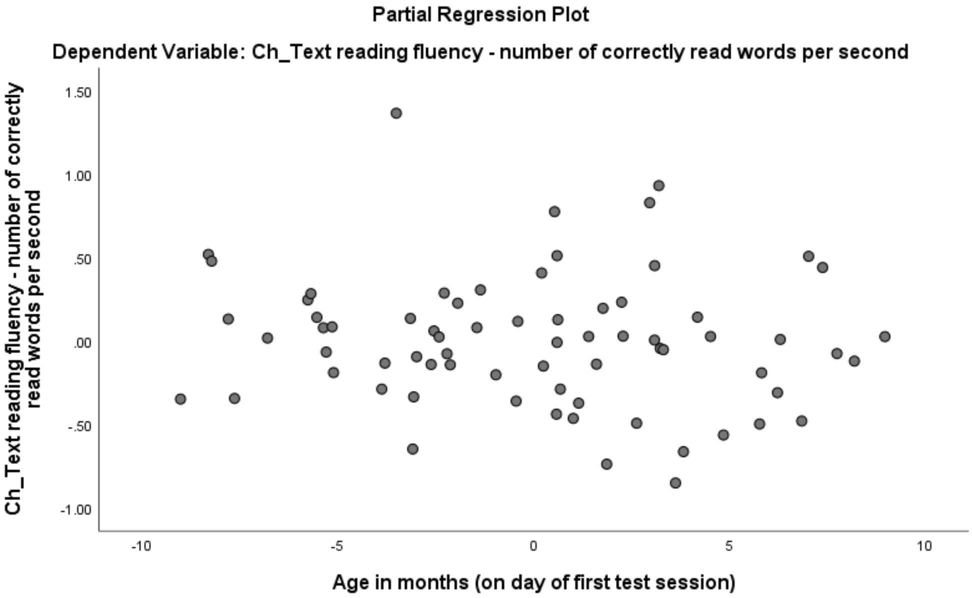

Figure A9. Partial Regression plot, age

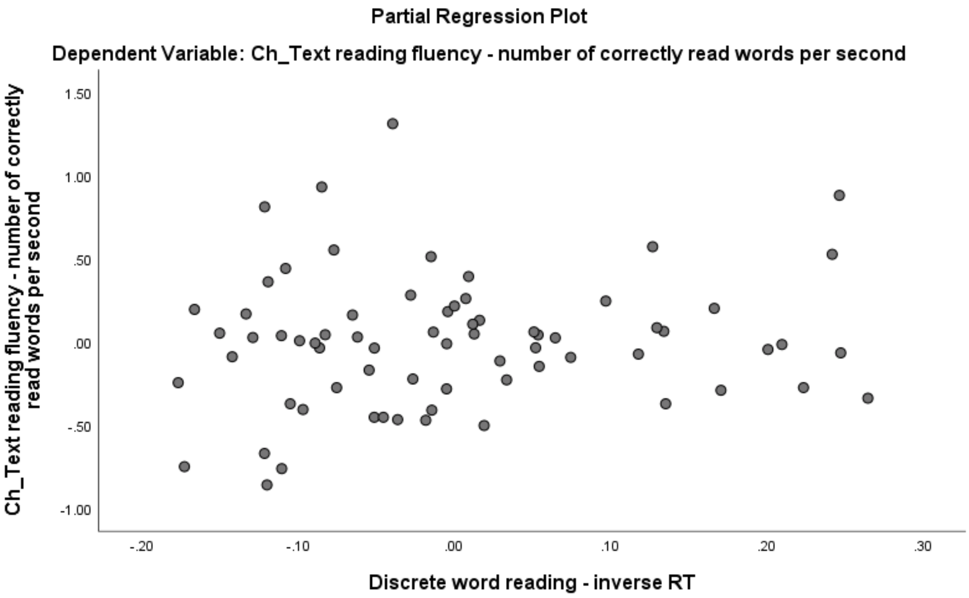

Figure A10. Partial Regression Plot, discrete WR

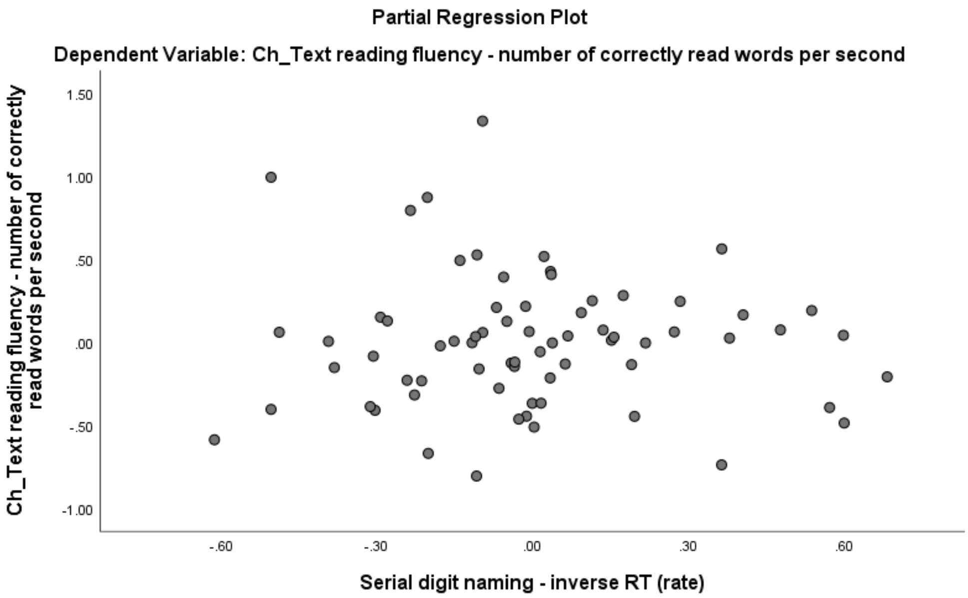

Figure A11. Partial Regression Plot, serial digit naming

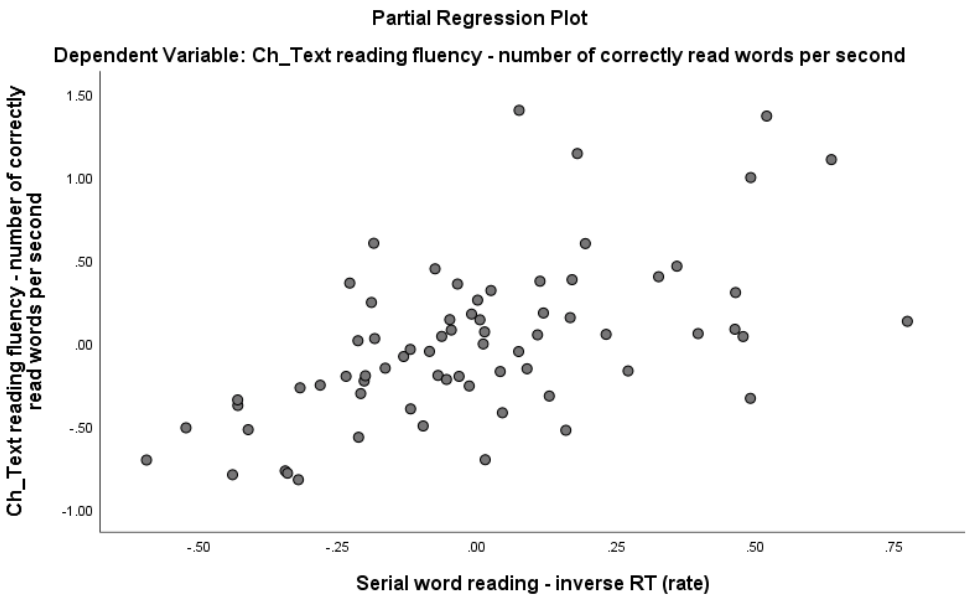

Figure A12. Partial Regression Plot, serial word reading

Table A1

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.115	1	.115	.406	.526[b]
	Residual	19.205	68	.282		
	Total	19.320	69			
2	Regression	8.371	4	2.093	12.424	.000[c]
	Residual	10.949	65	.168		
	Total	19.320	69			
3	Regression	8.570	5	1.714	10.205	.000[d]
	Residual	10.750	64	.168		
	Total	19.320	69			

a. Dependent Variable: Ch.Text reading fluency - number of correctly read words per second

b. Predictors: (Constant), Age in months (on day of first test session)

c. Predictors: (Constant), Age in months (on day of first test session), Serial digit naming - inverse RT (rate), Discrete word reading - inverse RT, Serial word reading - inverse RT (rate)

d. Predictors: (Constant), Age in months (on day of first test session), Serial digit naming - inverse RT (rate), Discrete word reading - inverse RT, Serial word reading - inverse RT (rate), PPVT NL - raw score (cutoff item minus number of errors, incl. cutoff set)

Appendix 2, Text Reading Fluency Task

Knuffelapen

De leestaak met opdrachten gaat beginnen. Als eerste opdracht pak je het gele blokje. Goed gedaan. Let op, je hoeft niet altijd de blokken te gebruiken. Aan het einde van de zin bijvoorbeeld spreek je het woord 'banaan' hardop uit. Heel goed. We houden bij of je dat hebt gedaan, dus je mag het woord nog een keer herhalen. Zeg het maar. Goed gedaan!

We gaan weer verder met lezen. Wist je dat er apen zijn die ook een beetje kunnen lezen? Niet echt natuurlijk, maar er zijn tekens die ze begrijpen. Weer wat geleerd! Om te laten zien dat je dit snapt, zwaai je nu even. Apen maken vaker testjes. Zo wilden mensen onderzoeken of knuffelen belangrijk is voor dieren. Daarom maakten ze twee namaak-moederapen die melk konden geven. Eén aap was gemaakt van zachte stof, en de andere van hard kippengaas. Ze lieten jonge aapjes kiezen tussen de twee moeders, en wat bleek? Dat vertel ik zo, maar klap eerst één keer in je handen. Goed zo!

Ik ga verder. Wat bleek? Bij een moederaap van gaas willen kleine aapjes niet drinken. Ook niet als ze heel veel honger hebben. Knuffelen lijkt dus net zo belangrijk als eten. Het is misschien zelfs iets belangrijker. Als je in de dierentuin oplet, zie je wel dat dit klopt. Apen zitten vaak aan elkaar te plukken. De tekst is nu bijna afgelopen. Je bent pas echt klaar als je een rood en een geel blokje pakt.

Translation: Cuddly monkeys

The reading task with assignments will start now. Your first assignment is to grab the yellow cube. Well done. Please note, you don't always have to use the cubes. At the end of this sentence you will say the word 'banana' out loud for example. Very well. We keep track of whether you did that, so you can repeat the word once again now. Just say it. Well done!

We continue with reading again. Did you know that there are monkeys that can also read a bit? Not really of course, but there are signs that they understand. Learnt something again! To show that you understood this, just wave. Monkeys have performed tasks before. At some point people wanted to assess whether cuddling is important for animals. That is why they made two fake monkey mothers that could give milk. One monkey was make of a furry fabric, and the other of rough metal wire. They let young monkeys choose between the two mothers, and what do you think? I will tell you that in an instant, but first clap your hands one time. Well done!

I continue. Guess what? The young monkeys didn’t want to drink from the mother that was made of metal wire. Not even when they were very hungry. Cuddling thus appears to be just as important as eating. It is maybe even somewhat more important. If you pay attention when in the zoo, you can see that this is often true. Monkeys tend to search through each other’s fur at lot. The text is now almost finished. You are really done when you grab a red and a yellow cube.