Rapid Naming, Reading- and Arithmetic Fluency

Pre-School Rapid Automatized Naming as a Predictor of Third-Grade Reading Fluency and Arithmetic Fluency

David Armas Hoff

Master's Thesis in Special Needs Education

Department of Special Needs Education

Faculty of Educational Sciences

UNIVERSITY OF OSLO

Spring 2021

Rapid Naming, Reading- and Arithmetic Fluency

Pre-School Rapid Automatized Naming as a Predictor of Third Grade Reading Fluency and Arithmetic Fluency

2021

Rapid Naming, Reading- and Arithmetic Fluency

Pre-School Rapid Automatized Naming as a Predictor of Third Grade Reading Fluency and Arithmetic Fluency

David Armas Hoff

http://www.duo.uio.no

Abstract

Background and rationale

In recent years, research on literacy has shown a particular interest in the topic of fluency. Building fluency seems to be an essential part of the transition towards skilled reading. Similarly, the concept of fluency has received attention in the realm of numeracy research. Here, findings suggest that arithmetic fluency might be an important facet of skilled mathematical performance. Further, reading fluency and arithmetic fluency overlap, suggesting that development may depend on some of the same abilities. Rapid automatized naming (RAN) is considered one of the best predictors of reading fluency and has also been found to predict arithmetic fluency. Combining these findings, it would be interesting to see whether the prediction of RAN extends to the overlap of reading- and arithmetic fluency.

In this study I will answer two research questions. The first question asks whether RAN predicts fluent performance in either of these academic domains. The second question asks whether RAN predicts common causes of the overlap between these types of fluency. The questions are:

> *Does RAN predict either reading fluency or arithmetic fluency, separately, after controlling for working memory and processing speed?*
>
> *Does RAN predict the covariance of reading fluency and arithmetic fluency, after controlling for working memory and processing speed?*

Method

This thesis is written in conjunction with the ongoing, longitudinal research project NumLit: Development of Numeracy and Literacy in children. NumLit is a project carried out by the Department of Education and the Department of Special Needs Education at the University of Oslo. For the final analyses, I use data obtained from 219 of the children participating in the project. These children have been tested annually since pre-school, on a range of measures relating to numeracy, literacy, as well as general cognitive abilities and language. The current study uses data from the first-, and the most recent wave of measurement; pre-school and third grade. The predictors: RAN, working memory, and processing speed, were measured in

pre-school. While outcome measures of reading fluency and arithmetic fluency, were measured in third grade.

Analyses

To answer the research questions, we will first attend to descriptive statistics and bivariate correlations. After this, structural equation modelling (SEM) is used to create latent factors and specify relationships between the variables. This allows us to go beyond prediction of performance on single measures, and instead predict performance in more holistic constructs of fluency. SEM analysis was carried out in R-Studio (RStudio Team, 2021), with the lavaan package (Rosseel, 2018). Descriptive statistics and bivariate correlations were provided through jamovi (The jamovi project, 2021).

Results and conclusion

Rapid automatized naming (RAN) provided a sizeable prediction of both reading fluency and arithmetic fluency, even after controlling for working memory and processing speed. When modelling the covariance of reading- and arithmetic fluency, as a second-order factor, RAN was the sole predictor of this factor. The prediction of RAN on the overlap of fluency, was found not to be due to demands for processing speed or working memory present in RAN performance. Processing speed instead predicted reading fluency directly, and working memory had a direct effect on arithmetic fluency. This means that processes involved in kindergarten RAN, are similar to processes common between the fluency outcomes in the third grade.

Preface

First and foremost, thank you to all the children who have participated in the NumLit project. Also, to the NumLit team, thank you for letting me take part in the project and meet the wonderful children behind the numbers and graphs.

I am immensely grateful to my supervisors, Tonje Amland and Athanassios Protopapas, for your guidance, thoughts and encouragement. Your enthusiasm for your research field is infectious and inspirational.

A big thanks to my family, and in particular my mother, for being kind, loquacious and supportive beyond belief.

Above all, my wife Sara and my son Leo, thank you for being my strength of will, my motivation, and for bearing with me in this prolonged period of academically induced partial disconnectedness from reality. All my love and pride belong to you.

Contents

1 Introduction ... 1
1.1 Background and rationale ... 1
1.2 Research questions ... 2
1.3 Structure of the thesis ... 2
2 Reading Fluency ... 3
2.1 From accuracy to fluency ... 3
2.2 Unitisation of the reading process ... 3
2.3 Sight-word reading and the lexical quality hypothesis ... 4
2.4 Definitions of reading fluency ... 5
2.5 Seriality and cascaded processing ... 6
2.6 The role of orthography ... 7
3 Arithmetic fluency ... 9
3.1 Importance of fluency ... 9
3.2 Arithmetic Strategies ... 10
3.3 Definition of arithmetic fluency ... 11
3.4 Processing of numbers – the triple-code model ... 11
3.5 The shift in strategy and Siegler's overlapping waves ... 13
3.6 What causes differences in strategy use ... 14
3.7 Development of strategy choice: The adaptive strategy choice model ... 15
3.7.1 Differences in strategy use for different operations ... 15
3.8 The combined picture ... 16
4 Rapid automatized naming ... 18
4.1 What is RAN? ... 18
4.2 Alphanumeric versus non-alphanumeric RAN ... 18
4.3 What does RAN measure? ... 19
5 RAN and Reading ... 20
5.1 RAN and reading fluency ... 20
5.2 What explains the relationship between RAN and reading? ... 21
5.2.1 Phonological processing ... 21
5.2.2 Orthographic processing ... 21
5.2.3 Orchestration ... 22
5.2.4 Domain-general abilities ... 22
5.2.5 Serial processing ... 22
5.3 RAN and Orthography ... 23
6 RAN and arithmetic fluency ... 24
6.1 Alphanumeric versus non-alphanumeric ... 24

6.2 Explanations for the association 25
6.2.1 Phonological processing 25
6.2.2 Serial processing 25
6.2.3 Domain-general abilities 25
6.3 Inconsistent outcome-measures 26
7 Previous research on the relationship between RAN, reading- and arithmetic fluency 27
7.1 Related abilities 27
7.2 Does the prediction of RAN change over time? 27
7.3 Is RAN a unique predictor 28
7.4 The role of time and developmental stages 29
7.5 Does prediction depend on format of RAN? 30
7.6 Explaining the three-way-relation 30
7.6.1 Similarities in strategy development - from procedure to retrieval 30
7.6.2 Similar reliance on a mechanism of visual-verbal associations 31
7.6.3 Similar reliance on retrieval of phonological representations 31
7.6.4 Cascaded processing 32
8 Present study 33
9 Methods 34
9.1 Design 34
9.2 Sample 34
9.3 Data collection 35
9.4 Analysis 35
9.5 Variables and measurement tools 36
9.6 Measures 37
9.6.1 Rapid automatized naming (pre-school) 37
9.6.2 Processing speed (pre-school) 37
9.6.3 Working memory (pre-school) 37
9.6.4 TOWRE word-reading fluency (third grade) 38
9.6.5 TOWRE pseudo-word fluency (third grade) 38
9.6.6 Oral reading fluency (third grade) 38
9.6.7 Measures of arithmetic fluency (third grade) 38
9.6.8 TOBANS addition and subtraction fluency 39
9.6.9 TOBANS addition and subtraction with carry 39
9.7 Validity 39
9.7.1 Internal validity 39
9.7.2 External validity 40
9.7.3 Construct validity 41
9.7.4 Statistical conclusion validity 41
9.8 Ethical concerns 43
10 Results 45
10.1 Descriptive statistics of variables 45

10.1.1 Transformation of variables ... 46
10.1.2 RAN objects ... 47
10.1.3 RAN colours ... 48
10.1.4 Processing Speed ... 48
10.1.5 Backward digit-span ... 49
10.1.6 Test of word reading fluency - words ... 50
10.1.7 Test of word reading fluency - pseudo-words ... 50
10.1.8 Oral reading fluency ... 51
10.1.9 TOBANS ... 51
10.1.10 TOBANS addition ... 52
10.1.11 TOBANS addition with carry – 3rd grade ... 52
10.1.12 TOBANS subtraction ... 53
10.1.13 TOBANS subtraction with carry ... 54
10.2 Bivariate correlations ... 55
10.3 Structural equation modelling ... 58
10.3.1 Indices of global fit ... 58
10.4 Specifying a structural regression model ... 59
10.4.1 The measurement model ... 59
10.4.2 Modelling the first research question – RAN as a predictor of fluency ... 59
10.4.3 Modelling the second research question – RAN as a predictor of overlapping fluency 60
10.5 Model of fluency ... 63
11 Discussion ... 65
11.1 The first research question ... 65
11.1.1 Why does RAN predict fluency? ... 65
11.1.2 Previous research ... 66
11.2 The second research question ... 66
11.2.1 Previous research ... 67
11.3 Results in a validity context ... 68
11.3.1 Internal consistency ... 68
11.3.2 External validity ... 68
11.3.3 Statistical conclusion validity ... 69
11.3.4 Construct validity ... 70
12 Conclusion ... 72
12.1 Implications for Special Needs Education ... 72
12.2 Limitations ... 73
13 References ... 74
14 Appendix ... 82

List of Tables

Table 1: List of Measures ... 36
Table 2: Descriptive statistics of the transformed variables ... 54

Table 3: Bivariate correlations between all variables 57
Table 4: Parameter estimates of the covariance model..87
Table 5: Correlation residuals of the covariance model 88

List of Figures

Figure 1: RAN Objects 47
Figure 2: RAN colours..48
Figure 3: Processing Speed...48
Figure 4: Backward digit-span..49
Figure 5: TOWRE words..50
Figure 6: TOWRE pseudo-words 50
Figure 7: Oral reading fluency..51
Figure 8: TOBANS addition..52
Figure 9: TOBANS addition with carry...52
Figure 10: TOBANS subtraction 53
Figure 11: TOBANS subtraction with carry..54
Figure 12: Model of overlapping fluency 62
Figures 13-68: Bivariate Scatterplots..82

Word count: 22947

1 Introduction

1.1 Background and rationale

Literacy and numeracy are essential abilities. Together, these skills provide a solid foundation for life-long learning. Being able to read and do calculations, is important for general quality of life as well as educational- and employment prospects (OECD & Statistics Canada, 2011). Attaining a functional level of efficiency in both skills, is therefore highly beneficial.

Interestingly, reading- and math abilities are related (Balhinez & Shaul, 2019; Koponen et al., 2020; Purpura et al., 2011). This also extends to a high co-occurrence of reading- and math difficulties (Willcutt et al., 2013). Why they overlap is not fully known. Predicting efficient performance in both these domains might offer insight into what causes the relationship (Cirino et al., 2018; Koponen et al., 2020). This efficiency - often described as fluency - is important in both domains. For reading, becoming fluent is regarded as an integral part of becoming a skilled reader (Schwanenflugel & Knapp, 2016). In a similar vein, fluent performance in simple arithmetic has been found to predict later mathematical achievement (Fuchs et al., 2016). If fluent performance is at the heart of the literacy-numeracy relationship, then finding cognitive predictors of this overlap might explain why these abilities are connected.

Rapid automatized naming (RAN) is a candidate that may provide such prediction. RAN tasks have been used extensively in reading research as a predictor of reading (Georgiou et al., 2009). In recent years RAN has also been used to predict mathematical ability (Koponen et al., 2017). However, RAN is a complex predictor, as RAN performance involves several subcomponents (Georgiou & Parrila, 2020). This means that RAN cannot pinpoint the exact processes that are common between fluent reading and arithmetic. For this purpose, using control variables such as working memory and processing speed can further illuminate the predictive nature of RAN.

1.2 Research questions

The present study aims to answer two research questions:

> *Does RAN predict either reading fluency or arithmetic fluency, separately, after controlling for working memory and processing speed?*

> *Does RAN predict the covariance of reading fluency and arithmetic fluency, after controlling for working memory and processing speed?*

1.3 Structure of the thesis

Chapters 2 through 7 make up the theoretical background section of the thesis. First, chapters 2 and 3 outline reading fluency and arithmetic fluency respectively. Here, the focus is on the development of strategies and important components underlying fluent performance. After this, chapter 4 describes what constitutes rapid automatized naming (RAN) tasks. The following two chapters each describe how RAN relates to reading fluency, and to arithmetic fluency. The seventh chapter then expands upon this by considering the overlapping relations of RAN, reading- and arithmetic fluency. The final chapter of the theoretical background section outlines the goals of the present study.

Chapter 9 is the methods section of the thesis. This chapter outlines the method employed to answer the research questions. Features of the study; the design, the statistical analyses employed, and measures are described. In addition, different aspects of validity, relevant for the study are discussed in section 9.7. Ethical concerns are also discussed in the last section of this chapter.

In chapter 10, the results of the study are provided. Descriptive statistics and bivariate correlations are provided, before specification of structural equation models is described.

The penultimate chapter discusses the results in light of theory. Here, the results are tied together to provide answers to each of the research questions. After this, the results are interpreted in a validity context, a section which also outlines limitations of the study. Implications for special needs education are then discussed, before a conclusion is provided.

Conclusions of the study are provided in chapter 12, together with implications for special needs education and a short synopsis of the limitations of the current study.

2 Reading Fluency

2.1 From accuracy to fluency

Learning to read is a complicated endeavour. Letters and written words are a meaningless jumble until children gain intimate knowledge of the system that they represent. Learning the alphabetic principle – that letters represent speech sounds - is key to learning how to read (Cain, 2010; Ehri, 2014). More specifically, the smallest unit of speech that distinguishes meaning, is known as a phoneme, and graphemes are written representations of phonemes (Schwanenflugel & Knapp, 2016). Understanding the alphabetic principle and learning the specific grapheme-phoneme mappings of the orthography in question, provides a fundamental basis for literacy. For a beginning reader, reading a word involves identifying letters and retrieving phonemes that match the graphemes in the word (Ehri, 2014). Thus, early literacy development hinges on accuracy.

Beginning readers may struggle with several aspects of this basic reading yet are expected to be accurate after a relatively short amount of time, given the apparent complexity of the task. Still, learning to read with accuracy is not enough. Skilled readers are expected to master several aspects of reading, with comprehension being the overarching goal. Expectations rapidly change and the metric of skilled reading shifts towards fluency.

2.2 Unitisation of the reading process

Becoming a fluent reader is thought to involve a shift from phonological recoding of letters, to consolidation- and reading of larger orthographic units (Cain, 2010; Ehri, 2014; Schwanenflugel & Knapp, 2016). In this way, sufficient accuracy is important both as a prerequisite for, but also as an integral part of reading fluency (Schwanenflugel & Knapp, 2016). Accurate phonological recoding functions as a sort of self-teaching strategy, allowing fluency to develop (Vaessen & Blomert, 2010).

Through practice, children become attuned to words and patterns in text, resulting in word-level knowledge that can be applied to lessen the demands of reading (Ehri, 2014). This process is described as consolidation or unitisation. There is a shift in the size of units that are processed in reading, from smaller to larger units (Ehri, 2014). Unitisation happens through orthographic mapping; pronunciations are linked with spelling patterns, allowing an increase in unit size (Ehri, 2014). The common spelling patterns of syllables and morphemes may for instance be used to access pronunciations of words, more rapidly than they would be accessed through grapho-phonemic connections (Ehri, 2014). This gradual build-up of word-level knowledge results in fluent reading through the use of units of increasing size, even encompassing whole words.

2.3 Sight-word reading and the lexical quality hypothesis

There are four main strategies for the reading of words: Phonological recoding, prediction, by analogy, or by sight (Cain, 2010; Ehri, 2014). The most basic of these is phonological recoding, where the phonemes associated with individual graphemes are produced and then blended in order to produce the pronunciation of the spoken word. Prediction involves a guess informed by context and superficial features of the word. Reading by analogy describes using other similar, known spellings to access pronunciations of a target word.

Finally, sight-word reading is the near-instantaneous processing of a whole word, and is considered to be a result of sufficient exposure to a word in its written form, allowing quick lexical access (Ehri, 2014). The frequency of sight-word reading is thought to underlie the transition towards fluent reading (Cain, 2010). This does not mean that sight-word reading fully replaces the use of other strategies, but rather that it becomes the predominant strategy given time and practice (Vaessen & Blomert, 2010).

Efficiency in word-reading and the prevalence of sight-word reading can be considered a result of lexical quality. Perfetti (2007) describes how features of the mental representations of words can differ, and how these differences explain the efficiency of retrieval: Sight word-reading is linked to how individual words are stored in the mind through representations of phonology, semantics and orthography, in what is called the mental lexicon. Each of these representations can vary in their quality as well as in their level of connectedness; the strength of their bonds.

For a beginning reader, mental representations of the word “dog” can be well established in terms of phonological and semantic representations. However, spellings such as “dogg” or “dawg” could easily be considered correct, due to the orthographic representation of the word “dog” not being fully specified or constant; parts of the orthographic representation are still variable.

Phonology and orthography are both representations of word form, and these representations must be well-specified to enable quick access (Perfetti, 2007). Well-specified representations of words with strong connections lead to automaticity and efficiency. Efficient reading is effortless, and the level of effort can be directly linked to the lexical quality of the relevant words in the reader's mind (Perfetti, 2007). Skilled readers have higher quality lexical representations on average than poor readers, but lexical quality is associated with individual words, meaning even skilled readers' lexical representations are of differing quality.

Lexical quality determining the efficiency of word-reading is also consistent with eye-movement research, where fixation time has been found to be associated with both the frequency and familiarity of words (Rayner & Slattery, 2009). Essentially, if a word is familiar the reader spends less time reading it.

2.4 Definitions of reading fluency

The main goal of reading is to be able to comprehend text. While the ability to accurately produce spoken words from text is quite useful, it is not sufficient for comprehension. A range of processes related to attention, visual identification, eye-movements as well as linguistic processes, contribute to reading both at the level of words and connected text (Norton & Wolf, 2012).

A common distinction is made between lower- and higher-level processes of reading. Lower-level processes are those that contribute to skilled word-level reading, and higher-level processes are those that enable reading comprehension (Schwanenflugel & Knapp, 2016). Reading of individual words is a complex process which requires the synchronised effort of multiple components. If one part of the orchestra is not on point, then the collective performance suffers. Lower-level processes are involved in the development of unitisation and automaticity, which in turn leads to higher-level processes becoming more engaged, and lower-level processes less so (Kuhn et al., 2010).

For skilled readers, reading has become autonomous; the underlying processes are automatized to the extent that virtually no conscious effort is required for reading to take place (Kuhn et al., 2010). Theoretically, resources can now be allocated to "higher-level" processes of comprehension (Kim, 2015; Schwanenflugel & Knapp, 2016). So, fluency is necessary, but not sufficient for reading comprehension. Instead, the reading process seems to become increasingly facilitated by qualities of connected text (Kim, 2015).

Definitions of reading fluency tend to emphasise the role fluency plays in skilled reading. Most definitions emphasise that fluency comes from a union of accuracy and rate (Kuhn et al., 2010; Schwanenflugel & Knapp, 2016). Others stress a need for comprehension and prosody as necessary for reading fluency (Norton & Wolf, 2012). This distinction concerns whether fluency is primarily separate from higher-level processes of comprehension, or whether fluency and comprehension are intrinsically linked (Kim, 2015; Priya & Wagner, 2009).

The view where comprehension is more involved in reading fluency, describes fluency not only in terms of the automaticity of word-level reading, but also as influenced by processes of comprehension (Norton & Wolf, 2012). Norton & Wolf (2012) use the term “fluent comprehension” to expand the concept of reading fluency, and to imply that the relationship between fluency and comprehension is bidirectional. In other words, reading fluently makes you comprehend more, but good comprehension also fosters more fluent reading.

Indeed, certain findings suggest that over time the fluent reading of connected text develops beyond the reading of unrelated words, and that text-reading fluency has a bidirectional relationship with reading comprehension (Kim, 2015). Here, fluency is a product not only of abilities relating to word-level reading, but also of separable skills related to text-level reading. One might say that this is not an alternate view of what reading fluency is, but rather of the role it plays in functional literacy.

2.5 Seriality and cascaded processing

The assumption that lexical quality and the frequency of sight-word reading is sufficient to explain reading fluency, has been questioned in recent years (e.g. Protopapas et al., 2018).

Altani et al. (2020) studied concurrent differences in reading among 710 English and Greek first-, third- and fifth-graders. They did this by matching tasks between “discrete” and “serial” conditions. In the discrete condition, a word-reading task would entail reading individually presented words on a screen, that changed once a response was uttered. In the serial condition, all words were instead simultaneously available in an array. Altani et al. (2020) found that the relationship between discrete and serial word reading was strongest in the early grades. However, while there was significant growth in serial reading performance in higher grades, discrete word reading performance seemed to stagnate.

If the efficiency of reading single words accounts for the development of reading fluency (i.e. sight-word reading), then discrete word reading across all skill-groups should be closely related to serial word reading, which it was not. An alternate explanation for this relationship is that serial word reading relies on simultaneous, parallel processing of multiple words (Protopapas et al., 2013). This is thought to entail a “buffering” where words are simultaneously processed at different stages of the reading process. While one word is being uttered, others are being processed or viewed.

This view, regarding a "cascaded" processing of multiple items, is not an alternative to sight-word reading, but rather a complement, to explain development of fluent reading past the process of unitisation. Sight-word reading represents the strategy of least constraint in terms of individual word reading, and cascaded processing expands fluency to deal with processing of word-sequences.

2.6 The role of orthography

Written alphabetic languages exist on a continuum from transparent to opaque (Share, 2008). This terminology describes the consistency of the phoneme-grapheme mappings of the orthography (Landerl et al., 2019). Transparent orthographies such as Finnish or Spanish represent the phonemic structure of spoken language consistently, which in turn makes it easier for beginner readers to read accurately through phonological recoding (Koponen et al., 2020; Korpipää et al., 2017). Opaque orthographies such as English, require the reader to be aware of a high number of irregularities, often due to the orthography representing both phonology and morphological units; units of meaning (Landerl et al., 2019).

English has a higher proportion of irregular words and inconsistent spelling patterns than more transparent orthographies, meaning many words cannot be read correctly by relying on phonological recoding (Cain, 2010). This may lead to different developmental trajectories (Share, 2008). For instance, while Finnish children tend to reach a ceiling of reading accuracy by the end of first grade, children learning the English orthography tend to struggle with accuracy for much longer (Koponen et al., 2020).

In research conducted in the English orthography, accuracy measures are often employed as these capture variability in skill for longer (Share, 2008). In transparent orthographies it is more common to use fluency measures, since accuracy measures quickly fail to distinguish between children's reading (Araújo et al., 2014; Share, 2008). Share (2008) points out that most literacy research is conducted in English-speaking countries, and thus through sheer volume, English has provided the de facto framework for understanding literacy development across orthographies, despite the English orthography arguably being an outlier.

Features of orthography could also determine the efficiency of strategies and which cognitive abilities are more engaged in reading (Park & Uno, 2015). For instance, phonological awareness has been found to be more predictive of reading outcomes in English than in transparent orthographies (Georgiou et al., 2008; Landerl et al., 2019). The psycholinguistic grain size theory suggests that features of orthography determine the size of units that are processed in reading (Ziegler & Goswami, 2005). The gist of this theory is that if smaller linguistic units are less consistent, and there is a higher degree of consistency in larger units, then children might start utilising larger linguistic units such as rimes, syllables and whole words at an earlier stage.

Qualities of orthography might also affect atypical development. Deficits in accurate reading are often considered to be primary indicators of word-level reading difficulties in opaque orthographies, where in more transparent orthographies fluency is a bigger concern (Diamanti et al., 2018; Share, 2008).

The present study is conducted in Norway, and Norwegian is a semi-transparent orthography (Arnesen et al., 2017), meaning there is more consistency in grapheme-phoneme relations than in English, yet less than in languages such as Spanish or Finnish. In Norwegian some phonemes can be written using different graphemes, such as the /kj/ sound. This means that there is not a 1:1 relationship between graphemes and phonemes, yet non-transparent relations still follow quite regular patterns (Torkildsen et al., 2019). Norwegian children are expected to master accuracy of reading and writing earlier than their English-speaking counterparts, yet slower than those in more transparent orthographies.

3 Arithmetic fluency

3.1 Importance of fluency

Mathematical ability is important. Math is in the classroom, at the heart of technology, science and finance. The influence of math is ubiquitous. Still, there are many who struggle with math, a struggle detrimental to health and employment outcomes (Butterworth et al., 2015). The trajectories of math skills are stable from an early age, setting the stage for early weaknesses to impact academic achievement and quality of life (Fuchs et al., 2016).

In school, it is not only expected that children become proficient in arithmetic, but such mastery is a prerequisite for more advanced operations. Whether early arithmetic performance is effortless and automatic, predicts later mathematical ability (Fuchs et al., 2016). Such effortless processing, referred to as arithmetic fluency, seems to be a foundational skill in development of mathematical ability, as it allows for the implementation of advanced algorithms (Rinne et al., 2020), and contributes to understanding in word-problems (Wang et al., 2020). Further, in advanced math, arithmetic fluency enables quick access to intermediate answers in multi-step calculation and frees up cognitive resources (Carr & Alexeev, 2011; Koponen et al., 2016; Rinne et al., 2020). In other words, to understand more complex math, quick access to simple arithmetic facts might be highly beneficial.

Despite its essential nature, mathematical ability varies greatly. Even within a single classroom there may be differences in mathematical achievement equivalent to a seven-year gap in ability (Dowker, 2015). For some learners, meeting expectations of mathematical achievement is particularly difficult. Mathematical difficulties can result from different impairments, such as problems with representation of quantity, reasoning, or problems with the retrieval of arithmetic facts (Karagiannakis et al., 2014). Retrieval of facts has been implicated as a central part of arithmetic fluency, and fluency problems can possibly indicate severe mathematical difficulties (Geary et al., 2012; Jordan et al., 2003a; Vanbinst et al., 2015). While problems of a procedural nature; those that primarily revolve around how to carry out operations, are seemingly easier to remediate, problems tied to retrieval and arithmetic fluency are more stable (Chong & Siegel, 2008; Geary et al., 2012).

3.2 Arithmetic Strategies

Arithmetic ability is core to mathematical competence. Early strategies of single-digit arithmetic rely on counting techniques, and later strategies utilise fact retrieval either directly or "derived" through known facts (Dowker, 2014; Gilmore et al., 2018). Among these early strategies, those employed for addition are most described. These strategies of addition also provide a basis for the strategies of other operations, such as subtraction or multiplication (Dowker, 2014).

Gilmore et al. (2018) describe five common strategies for single-digit addition: The first arithmetic strategy children use is typically the "count all" strategy, where the child adds two numbers by counting up to their combined sum, starting from 1. Over time, children employ more efficient varieties of this strategy where they "count on", meaning they no longer start from 1, but from one of the numbers that are being added together.

The most rudimentary version of this strategy is "count on from first", where the child "counts on" from the first addend in the problem. A more efficient variety of this is to "count on from largest", where the child counts on from the larger addend in order to minimise the counting required to reach the answer. The use of this more advanced counting strategy is also considered by some to reflect that the child understands the concept of commutativity; that the order of the addends is irrelevant (Dowker, 2014; Gilmore et al., 2018).

Once children are sufficiently familiar with simple addition problems, their strategies become increasingly reliant on fact retrieval (Gilmore et al., 2018). Fact retrieval may be utilised either directly, or in a derived fashion (Dowker, 2014). Direct fact retrieval simply means that the child is sufficiently familiar with a problem and its solution, enabling them to retrieve the solution as an arithmetic fact from long-term memory (Gilmore et al., 2018; Kaufmann et al., 2004). Cowan et al. (2011) add that for fact retrieval to be efficient, it should be rooted in semantic understanding, not rote learning of phonological forms. In this way counting strategies provide a meaningful basis for learning arithmetic facts.

Derived- or decomposed facts is a strategy that utilises known facts to solve a problem for which direct retrieval of the solution is unavailable (Gilmore et al., 2018). For instance, when solving 9 + 3, if the child can retrieve related facts, they can quickly determine that 3 is the sum of 2 + 1, that 9 + 1 is 10, and that 10 + 2 is 12. This somewhat roundabout way utilises memorised facts in order to produce a solution to a problem without the need for counting.

Findings suggest that not all children transition to predominant use of fact retrieval, some continue to rely on slow and unreliable procedural strategies even for simple arithmetic (Gilmore et al., 2018; Jordan et al., 2003b; Vanbinst et al., 2015). Further, children with deficits in fact retrieval show lower overall mathematical achievement, and these deficits are seemingly persistent (Geary et al., 2012; Jordan et al., 2003b).

3.3 Definition of arithmetic fluency

Research on numeracy development is considered to lag behind reading research (Thomas, 1987). Consequently, there is less agreement on which aspects of development are most central. As such, definitions of arithmetic fluency are sparse relative to definitions of reading fluency.

Balhinez & Shaul (2019) describe automatized fact retrieval in single-digit arithmetic as the core-component of arithmetic fluency. Poor arithmetic fluency is, as mentioned above, thought to reflect reliance on less efficient, procedural strategies rather than retrieval (Vanbinst et al., 2015). This means that familiarity with problems, and problem-solution connections, are seen the key to automaticity (Lemaire & Siegler, 1995). One might interpret from this, that the development of early strategies largely determines later arithmetic fluency.

Other definitions of arithmetic fluency do not place the same emphasis on strategy and fact retrieval, but instead view arithmetic fluency simply as the rate with which children solve single-digit arithmetic (e.g., Carr & Alexeev, 2011). This does not rule out that fact retrieval plays an important role in automaticity. It simply reflects a broader, more pragmatic approach; a definition which also mirrors how arithmetic fluency is usually measured.

Arithmetic fluency is generally measured in terms of efficient addition and subtraction for young children, as multiplication and division are learnt, and become efficient, at a later point (Koponen et al., 2017). Multi-digit arithmetic relies on understanding of concepts such as place-value, monitoring and procedural processes, and performance in multi-digit arithmetic may therefore less effectively capture fact retrieval (Koponen et al., 2017).

3.4 Processing of numbers – the triple-code model

In order to understand fact retrieval and arithmetic fluency, we need to understand how numbers are represented in the mind. Dehaene’s (1992) triple-code model outlines three interconnected codes that make up the mental representations of numbers: The analogue magnitude code represents the semantic component of numbers, as activations on a mental number line. The visual Arabic code is responsible for the symbolic representations of Arabic digits. The auditory verbal code relates to phonological forms such as number words.

Each of these representations are thought to interface with input in their respective formats and be responsible for their associated outputs. For instance, the auditory verbal code is thought to be related to inputs and outputs in auditory form, as well as number words in written format, as these represent phonological information (Dehaene, 1992). Arithmetic facts also are hypothesised to be associated with the auditory code, meaning facts such as; 3 + 5 = 8 may be represented in a verbal frame as “three plus five is eight” (Koponen et al., 2007). This is in line with findings of phonological processing being associated with rate of fact retrieval (De Smedt & Boets, 2010).

Children’s ability to determine which of two digits represents the larger magnitude, known as symbolic magnitude comparison, has also been found to correlate positively with fact retrieval (Vanbinst et al., 2012, 2015). This suggests that the quality of symbolic number representations might be important for arithmetic fluency. Further, symbolic magnitude comparison skill predicts mathematical achievement beyond the contribution of non-symbolic magnitude comparison (Malone et al., 2019; Vanbinst et al., 2015). These results could indicate that once symbolic representations become available, their non-symbolic counterparts become less utilised. Perhaps the closeness of mapping between visual and semantic forms makes analogue semantic representations redundant. Alternately, mental arithmetic might simply become increasingly abstract and ceases to rely on analogue representations of magnitude. In any case, these findings suggest that the quality of representations, or of mappings between different codes, might be a gateway to fluent arithmetic.

These mappings between representations, and the strength of such connections between codes, may be of importance. Malone et al. (2019) investigated whether children’s ability to learn to associate magnitude representations with corresponding visual and auditory forms, was related to mathematical ability. The 166 first- and fourth-grade participants were tested on their ability to learn connections between non-symbolic magnitude stimuli (dots) and either visual symbols or non-words. They found that children’s ability to exhibit such paired-associate learning, predicted their arithmetic ability even after controlling for a range of early predictors of literacy and numerical skills.

The triple-code model represents a hypothesis that arithmetic performance cannot be separated from the format of problem. Studies show that format has a substantial effect on both accuracy and fluency of operations, with adults displaying 30% slower, and 30% less accurate performance when problems are presented in number-word form, rather than Arabic numeral format (Campbell, 1994, 2015). The format of problems also seem to affect which strategies are used, retrieval being 60% less likely when problems are in word format (Campbell, 2015). These findings suggest that format is important, consistent with the triple-code model.

3.5 The shift in strategy and Siegler's overlapping waves

If mental representations are the building blocks of arithmetic fluency, then strategy development is the building process. As already mentioned, strategies of arithmetic seem to progress from procedure towards retrieval. However, learning a new strategy does not mean less efficient strategies are discarded. In fact, fourth grade children (Foley et al., 2017), and adults (Polspoel et al., 2017) have been found to sometimes revert to counting strategies for single-digit arithmetic. Strategy development is seemingly a more fluid process than a process of distinct steps.

Siegler (1996) describes children's development of strategy usage as a series of overlapping waves, where each wave represents the frequency of a specific strategy. Early strategies such as "count all" are used for all problems for as long as it remains the child's only viable strategy. In the initial stages of learning a new, more efficient strategy such as "count on from first", the less effective strategy of "count all" is still likely be most frequently used.

Siegler (1996) goes on to present three aspects of importance for strategy choice: Speed, accuracy, and automaticity. In short, the child will choose among their available strategies the one they deem to be most efficient. The general tendency to move towards retrieval strategies is a natural consequence of these strategies generally being quicker and producing more accurate answers. However, it is still the proficiency of the individual that determines their use of strategy, rather than the general utility of the strategy.

Studies have shown that strategy repertoire is not necessarily different among those who exhibit low and high mathematics achievement, but those who are more skilled utilise more efficient strategies to a greater extent, and also adapt strategies depending on the problem (Gilmore et al., 2018). Individual preferences may also play a role. Some studies find a group of mathematically capable children dubbed "perfectionists", children who place an added emphasis on accuracy (Carr & Alexeev, 2011). Even if they retrieve the answer, these children may feel more comfortable double-checking this retrieved solution with a procedural strategy, to ensure that the answer is correct.

3.6 What causes differences in strategy use

The large variability in individual strategy use, has been hypothesised to stem from variability in a range of different cognitive factors. Imbo & Vandierendonck (2007) found that verbal working memory, as indexed by digit-span, was related to children’s use of arithmetic strategies. Further, they found that taxing working memory while children were solving addition problems, had a greater detrimental effect on the accuracy of procedural strategies than direct retrieval. This suggests that direct retrieval is less reliant on working memory resources. Further, children with higher working memory capacity might be more capable in their use of procedural strategies, which in turn allows them to transition to use of fact retrieval more quickly, freeing up working memory resources.

Foley et al. (2017) found that children’s use of decomposition strategies was associated with another aspect of working memory, namely their short-term visuospatial memory. They suggest that visuospatial memory could enable children to create a sketch for mapping calculations, simplifying the use of decomposition strategies. Thus, they claim that visuospatial memory is related to mathematical achievement through strategy use.

Koponen et al. (2016, 2020) argue that counting skills are predictors of arithmetic fluency and fact retrieval. Counting skills such as counting backward- or skip-counting enable higher accuracy in use of procedural strategies, and thus provide a basis for a shift to fact retrieval. If a child can accurately retrieve a chain of number words and is able to break apart these chains and manipulate them, this enables more advanced counting strategies. This means that counting skill may be related to fact retrieval through a common reliance on retrieving phonological representations, and also indirectly through contributing to procedural strategies.

Another finding that may explain differences in strategy use, is that of symbolic magnitude comparison being positively associated with fact retrieval (Vanbinst et al., 2012, 2015). This could suggest that symbolic representations provide a semantic context for the acquisition of facts. Thus, strategy development could be informed by the quality of - and access to - symbolic representations of numbers.

3.7 Development of strategy choice: The adaptive strategy choice model

Regardless of which cognitive abilities inform the use of strategy, there are likely some universal mechanisms that determine strategy use. Lemaire & Siegler's (1995) adaptive strategy choice model (ASCM) is a simulation model which operates under the assumption that strategies are picked based on their relative strength. Here, the strength of a strategy reflects three aspects of efficiency: The efficiency of the strategy for operations in general, for similar problems and for a specific problem. If a retrieval strategy is chosen, then a set of answers associated with the problem in question is scanned. The answer with the highest degree of association will then be chosen, but only stated if it also exceeds a confidence threshold. If no solution is provided through retrieval, a procedural strategy is employed.

In the ASCM, strategy choice is seen as a self-perpetuating process where the level of confidence leads to a solution being provided, further increasing the problem-solution association. When procedural strategies are used, association between problem and the correct solution is gradually built over time. Even if accuracy is low, no single wrong answer is likely to build higher association with the problem than the correct answer.

According to this model, more advanced strategies will naturally replace less sophisticated ones. While individual efficiency will vary, the choice of strategies will naturally tend towards those that are objectively more efficient. This tendency also predicts the use of procedural strategies to be more widespread for complicated problems: Fact retrieval in more complicated problems require more practice, as producing erroneous answers are more likely, and errors dilute association with the correct answer.

Models such as the ASCM also account for why erroneous answers tend not to be random, but rather related to one of the operands (Gilmore et al., 2018). A produced solution, whether correct or incorrect, will be increasingly associated with a problem. Over time, those answers that are incorrect, but related to the problem in some way, are more likely to be produced and associated with the problem.

3.7.1 Differences in strategy use for different operations

The ASCM describes how the perceived strength of strategies determines strategy selection. Convergent with this, research suggests that strategy is utilised flexibly for different types of operations (Siemann, 2018). Strategies for subtraction are considered to reflect the inverse varieties of addition strategies, requiring children to first understand the inverse relation of addition and subtraction (Dowker, 2014). However, fact retrieval is utilised much less in subtraction than in addition, despite the similar procedural strategies (Rinne et al., 2020).

One possible explanation for this is that subtraction problems are generally practiced less. According to the ASCM, associations between problem and solution may therefore not cross the required confidence threshold for fact retrieval.

For single-digit multiplication, fact retrieval seems to be more frequent than in both addition and subtraction (De Smedt, 2018; Rinne et al., 2020). A possible explanation is that since procedural strategies for multiplication are not very efficient, the confidence threshold for fact retrieval is lower. The inefficient nature of these procedural strategies might also be what prompts rote learning of multiplication tables to be a common teaching-strategy, further cementing retrieval as a preferred strategy.

3.8 The combined picture

Theoretical accounts of arithmetic fluency largely equate fluency with the frequency of fact retrieval. Thus, fluent arithmetic relies on problems to elicit strong associations between problem and solution. Alternately, fluency is also determined by the efficiency of procedural strategies.

The triple-code model presents a framework to understand how numbers are represented in the mind. In particular, the quality of symbolic number representations could determine the semantic context of arithmetic. Children with high-quality mental representations, could more easily be able to judge whether a solution makes sense, and more apt in their adaptation of advanced strategies. Conversely, poor representations could lead to error-prone execution which in turn leads to a stunted strategy development and lower rates of fluency.

There is a clear hierarchy of strategies, where retrieval is the most efficient. However, individual factors seemingly play a large part in strategy development. The strength of a strategy is a result of individual competence, which in turn may be informed by cognitive factors. The hierarchy of strategies describes which strategies have the potential for the highest level of fluency, yet the actual efficiency is the result of individual factors.

This all leads to an ultimate operationalisation of arithmetic fluency as the accuracy and speed of arithmetic for problems where retrieval is likely to be used (e.g. Koponen et al., 2017). As such, in the present study only simple addition and subtraction is considered, both with and without crossing the ten-boundary (i.e. “with carry”). Given the provided theoretical accounts, it makes sense to consider arithmetic fluency as largely tied to the frequency of retrieval. This also provides a meaningful basis for understanding the construct in relation to mathematical ability. If arithmetic fluency represents important aspects of strategy development and utilisation, then it can be interpreted as part of a causal pathway that connects cognitive factors to later mathematical achievement.

4 Rapid automatized naming

4.1 What is RAN?

Rapid automatized naming (RAN) tasks measure timed performance in naming a set of familiar items presented repeatedly in an array (Georgiou & Parrila, 2020). Commonly used stimuli are colours, pictures, digits or letters. Originally designed to determine features of alexia; reading failure related to brain damage, the test was also found to elicit poor performance in children with developmental dyslexia (Norton & Wolf, 2012).

RAN is considered to be one of the best predictors of reading ability regardless of orthography (Araújo et al., 2014; Landerl et al., 2019; Norton & Wolf, 2012). One explanation for this is that RAN represents a condensed version of reading, utilising many of the same underlying processes (Araújo et al., 2014; Norton & Wolf, 2012).

In recent years there has also been an increased interest in whether RAN is predictive of mathematical ability. Findings on the relationship indicate a moderate, positive association between RAN and mathematical abilities (Koponen et al., 2017). Further, this association seems to be higher for measures of arithmetic, in particular of arithmetic fluency, than for general math achievement (Koponen et al., 2017).

There are indications to suggest that growth in RAN ability is steep in the early primary school grades, and flattens over time (Wolf & Bowers, 1999). Further, according to Vander Stappen & Reybroeck (2018) RAN performance may be improved through intervention.

4.2 Alphanumeric versus non-alphanumeric RAN

Variations in the stimuli used in RAN, can be distinguished into two main formats: Alphanumeric RAN, comprised of letters or numbers; and non-alphanumeric RAN, consisting of any other visual stimuli, typically pictures or colours (Araújo et al., 2014; Hornung et al., 2017; Norton & Wolf, 2012). The choice of format can be influenced by factors such as the purposes of testing or the age of children, as younger children are unlikely to be familiar with letters and digits.

The two formats of RAN have been found to be differentially predictive of reading, as alphanumeric RAN tasks generally elicit higher associations with reading performance (Araújo et al., 2014; Lervåg & Hulme, 2009). This is not the case for arithmetic, where both formats are found to be similarly predictive (Koponen et al., 2017). Overall, there is a tendency for alphanumeric RAN to be processed more rapidly than non-alphanumeric RAN, likely due to alphanumeric stimuli representing smaller sets that more easily become automatized through over-learning (Wolf & Bowers, 1999).

4.3 What does RAN measure?

In their seminal work on the double-deficit theory of dyslexia, Wolf & Bowers (1999) critique the reigning assumption that RAN is simply an index of rate of access to phonological representations, and therefore simply an indicator of phonological processing. They argue that this practice does not sufficiently capture the processes that lead to the association between RAN and reading ability. They do not dispute that phonological processing is part of RAN performance. Instead they claim that more parts of RAN performance must be considered, in order to explain why the task predicts reading. Wolf & Bowers (1999) also describe a set of processes involved in letter-naming, to elucidate the complexity of RAN performance:

a) Attention to stimulus
b) Visual identification: where features of the stimuli are detected, discriminated and integrated in order to identify the letter
c) Visual features and patterns are integrated with orthographic representations
d) The visual information is integrated with phonological representations
e) Access and retrieval of phonological labels
f) Activation and integration of semantic & conceptual info
g) Motoric activation leading to articulation

To simplify this slightly: Attending to the stimulus is the basis for visual identification. In turn visual identification happens at several levels; at the feature level, for instance identifying the roundness of the letter O. This in turn leads to identification of the appropriate letter. This identity is the key to lexical access where the orthographic representation, or in this case the mental representation of a letter, activates its associated phonological representation; the phoneme, and a label; the letter name. The semantic and conceptual info in this case represents meaningful associations. Finally, all of this leads to the articulation of the letter name. Taken together, this represents an outline of the myriad of processes that underlie RAN performance.

5 RAN and Reading

RAN has been a long-standing staple predictor in reading research. However, the exact nature of its prediction on reading outcomes, is a matter of debate (Georgiou & Parrila, 2020). Where other measures can be more directly linked to distinct skills or abilities, the multicomponent nature of RAN means it is not necessarily indicative of any single ability (Norton & Wolf, 2012; Wolf & Bowers, 1999). What RAN indicates is therefore a controversial subject.

Some claim that RAN task performance indicates “naming ability” or simply “RAN-ability” (e.g., Balhinez & Shaul, 2019; Georgiou et al., 2020; Hornung et al., 2017; Korpipää et al., 2017). Others more cautiously categorize RAN as a task instead of an indicator of any specific ability (e.g., Norton & Wolf, 2012; Wolf & Bowers, 1999). As both reading and RAN involve a lot of different processes, explaining their relationship through any single aspect of RAN might prove insufficient (Norton & Wolf, 2012). Instead, a broader perspective may be required, where understanding combination and coordination of components involved in RAN is necessary to understand why it predicts reading.

In a meta-analysis on the relationship between RAN and reading, Araújo et al. (2014) find that RAN is related to word-reading, non-word reading as well as reading comprehension, and that the type of outcome measure influences the strength of the relationship. Overall, they found a moderate to strong relationship between RAN and reading ($r = .43$), and that RAN correlated more strongly with measures of reading fluency ($r = .49$), than with measures of reading accuracy ($r =.42$). Further, they describe that the relationship between RAN and reading accuracy seemingly decreases over time, while the relationship between RAN and reading fluency remains stable.

5.1 RAN and reading fluency

Interest in RAN has grown out of an increased interest in reading fluency as a facet of skilled reading. Among the cognitive factors found to predict reading ability, phonological awareness and RAN are considered perhaps the most important (Araújo et al., 2014). A common finding is that phonological awareness is most predictive of early reading, whereas RAN continues to predict reading across development (Vaessen & Blomert, 2010). This could insinuate that phonological awareness; the ability to identify and manipulate units of spoken language (Cain, 2010), is important for early reading, when phonological recoding is still an important strategy. However, RAN seems to be important at all stages, perhaps because it indexes fluent processing regardless of unit size. The contributions of phonological awareness and RAN may be affected by orthographic depth, as in transparent orthographies reading accuracy reaches a ceiling early in development (Landerl et al., 2019).

If phonological awareness is mainly important for accurate phonological recoding, then the relative contribution it makes to reading should decrease as lexical quality increases and sight-word reading becomes more prominent. However, RAN is seemingly a predictor of reading both when accuracy is a key component, but also later when accuracy is deemphasised (Araújo et al., 2014).

5.2 What explains the relationship between RAN and reading?

5.2.1 Phonological processing

Several hypotheses try to explain the observed relationship between RAN and reading. A commonly cited explanation is that RAN is a measure of rate of access to- and retrieval of phonological representations (Torgesen et al., 1997). This means that the observed relationship between RAN and reading is due to similar demands of phonological processing. Phonological processing describes processing that utilises mental representations of the sound-structures in spoken-language (Torgesen et al., 1997). Proponents of this explanation do not believe RAN is interchangeable with other measures of phonological processing; skills such as phonological awareness, but believe RAN is a member of a broader category of phonological processing measures.

One criticism of this explanation is that the demands of serial- and discrete naming are similar in terms of phonological processing, but that serial-naming predicts reading above and beyond the contribution of discrete-naming (Papadopoulos et al., 2016). Also, Vander Stappen & Reybroeck (2018) found that a phonological awareness intervention contributed only to PA performance and not to RAN, and vice versa; that a RAN intervention only benefited RAN performance, not phonological awareness.

5.2.2 Orthographic processing

Another explanation comes from Bowers & Wolf (1993) who claim that the relationship is due to naming speed, particularly of alphanumeric RAN, mediating processes that lead to unitisation of orthographic patterns. Becoming sensitized to common orthographic patterns, relies on a responsiveness to orthographic units. Problems with orthographic processing could start with a deficit in the visual identification of letters, which in turn slows development of pattern recognition and word-level knowledge (Papadopoulos et al., 2016). However, RAN predicts both the fluency of word- and pseudo-word reading similarly, two outcomes that rely differently on orthographic processing (Moll et al., 2009). This indicates that orthographic processing does not underlie the RAN-reading relation.

5.2.3 Orchestration

A third proposed explanation is that one cannot explain the RAN-reading relationship through specific parts of RAN, but rather that the relationship is due to similar demands for orchestration of several underlying processes (Norton & Wolf, 2012). Here, some parts of RAN might still be more salient in explaining the relationship to reading, but the overall predictive value of RAN comes from subprocesses' ability to function in unity. Support for this account may be garnered from the observation that a wealth of research has attempted to find mediators for the RAN-reading relation, but fail to find a mediator that explains away RAN's prediction of reading (Georgiou & Parrila, 2020).

5.2.4 Domain-general abilities

Others claim that the relationship is due to a common reliance on domain-general abilities such as processing speed. For instance Kail et al. (1994; 1999) claim that the association between RAN and reading is due to age-related increases in global processing speed. Other studies also report that general processing speed is implicated in RAN, but that it does not fully explain the relationship between RAN and reading (Georgiou et al., 2009; Vaessen et al., 2009).

5.2.5 Serial processing

Yet another explanation focuses on seriality as a component in both RAN and reading. Rapid automatized naming is more predictive of word reading in a serial format, than when words are presented individually (Protopapas et al., 2013). Also, discrete-RAN, where stimuli are presented individually, is less predictive of reading fluency than serial-RAN, where stimuli is presented simultaneously in an array (Altani et al., 2020). Here, the salient feature of RAN, underlying the relationship with reading fluency, is the fact that both of these utilise cascaded-processing; while one item is being uttered, following items are already being processed (Protopapas et al., 2013).

5.3 RAN and Orthography

The finding that RAN predicts the fluency- more than accuracy of reading, is consistent across orthographies (Araújo et al., 2014; Moll et al., 2014; Song et al., 2016). However, it is difficult to establish whether RAN relates differently to reading depending on orthography. As detailed in chapter 1, reading achievement tends to be measured differently across orthographies, and literacy instruction also proceeds differently. Making cross-linguistic, age-equivalent comparisons is therefore difficult, especially when considering that comparison requires test materials be tailored to match across languages, all while taking orthography into account (Landerl et al., 2019).

Altani et al. (2017) conducted a study with third-grade children speaking Greek, English, Korean and Chinese, each sample representing one of four widely different orthographies. They found that even after controlling for discrete naming of digits as well as discrete word reading, serial digit naming still predicted word reading fluency in all four samples. This indicates that the prediction of serial RAN taps at least partly the serial-processing aspect of reading, regardless of unit-size and orthography.

Moll et al. (2014) found that RAN was more predictive of reading fluency than reading accuracy across five different orthographies: English, French, German, Hungarian and Finnish. The only language where RAN significantly predicted reading accuracy, after controlling for predictors such as phonological awareness and phonological memory, was English.

6 RAN and arithmetic fluency

In a meta-analysis of the relationship between mathematical ability and RAN, Koponen et al. (2017) found a moderate correlation (r = .37). Correlations were also higher when timed measures were used, and when outcome measures used single-, rather than multi-digit calculations. The RAN-arithmetic relationship has been found to be unidirectional, as early RAN predicts later arithmetic fluency, but early arithmetic fluency does not predict later RAN (Georgiou et al., 2020).

RAN's unique prediction of variance in arithmetic fluency is found by a number of different studies (Koponen et al., 2016, 2020). Other studies fail to find a similar unique relationship, finding instead that RAN's prediction can be explained by domain general abilities such as general processing speed (Georgiou et al., 2013; Wang et al., 2020). This suggests that the relationship of RAN and arithmetic can be explained away by mediation, unlike the relationship of RAN and reading.

6.1 Alphanumeric versus non-alphanumeric

Where reading, and in particular reading difficulties are seemingly more highly correlated with alphanumeric RAN than with non-alphanumeric RAN (Araújo & Faísca, 2019), studies have found the relationship between RAN and mathematics to be similar for both RAN formats (Donker et al., 2016; Koponen et al., 2017). A possible explanation for this is that the target of alphanumeric RAN is available at a surface level, whereas non-alphanumeric stimuli might require additional retrieval of semantic knowledge (Donker et al., 2016). Essentially, a letter or a digit is closely associated with the name of the stimuli, whereas a coloured circle or a picture of an item is more closely associated with aspects of meaning. This suggests that the processing of non-alphanumeric RAN more closely resembles the processes of arithmetic, in that access to meaning is required in both.

Hornung et al. (2017) found that "quantitative numeral RAN", more specifically digits or finger-numeral-configurations, predicted unique variance in arithmetic fluency when controlling for an assortment of RAN types. Such findings indicate that something about stimuli related to numbers, predicts variance in arithmetic fluency beyond the general features of RAN. This is also in line with findings that suggest fingers not only provide concrete representations used in counting strategies, but are also involved in mental representations of arithmetic (Andres & Pesenti, 2015). Similarly, the prediction of digit-

RAN may suggest that the quality of mental representations of numbers, may affect both rapid naming and fluent arithmetic.

6.2 Explanations for the association

A number of explanations have been proposed for the relationship between RAN measures and arithmetic. Some of these mirror explanations of the relationship between RAN and reading, albeit adapted to the processes involved in arithmetic. It is also worth noting that due to the multifaceted nature of arithmetic, the prediction of RAN could vary widely for different types of arithmetic operations (Koponen et al., 2017).

6.2.1 Phonological processing

One account claims that RAN and fluent arithmetic both depend on access to - and retrieval of - phonological representations from long-term memory (De Smedt, 2018; Koponen et al., 2016, 2017). As mentioned in section 3.4, there is an assumption that fact retrieval relies on phonological representations. This is also convergent with phonological awareness seemingly being involved in fact retrieval (De Smedt, 2018; De Smedt & Boets, 2010) Further, this could mean that the prediction of RAN on arithmetic outcomes, may mainly depend on how frequently retrieval is used.

6.2.2 Serial processing

Koponen (2020) claimed that counting and RAN predict arithmetic due to similarities in the serial nature of retrieval. They propose that counting is associated with RAN due to similar demands for cascaded processing. Serial retrieval of phonological forms is shared across counting, RAN, and arithmetic. Such measures also share demands for attending to serially presented stimuli. This is similar to the previously described phonological processing-based explanation, but takes this a step further and implies that some of the variance RAN explains in arithmetic fluency, is due to similar demands placed on serial processing of items in sequence.

6.2.3 Domain-general abilities

Another explanation is that the relationship is due to common reliance on domain-general abilities such as general processing speed or working memory (Georgiou et al., 2013; Wang et al., 2020). Wang et al. (2020) found that the relationship between RAN and arithmetic fluency was no longer significant when controlling for processing speed and inhibition, meaning the effects of RAN on arithmetic fluency were mostly indirect. Similarly, Georgiou et al. (2013) found that an observed relationship between RAN and arithmetic could be explained by processing speed, phonological awareness and phonological short-term memory, whereas the same set of covariates fail to explain RAN's relationship with reading.

6.3 Inconsistent outcome-measures

Differences in the observed relationship between RAN and arithmetic fluency could potentially be due to differences in outcome measures. Measures of arithmetic fluency vary widely in the content they cover, most using some form of simple addition and subtraction, but others also opt to add multiplication (e.g. Rinne et al., 2020), and even division (e.g. Korpipää et al., 2017). Measures might be conceived to capture a wide span of arithmetic performance and to fit the age-group, yet an important question is what constructs they are thought to measure. If arithmetic fluency is synonymous with the relative frequency of retrieval, then it only makes sense to include problems that participants of the sample in question are likely to solve by use of retrieval.

7 Previous research on the relationship between RAN, reading- and arithmetic fluency

7.1 Related abilities

Findings suggest that early skills in literacy and mathematics are related (Balhinez & Shaul, 2019; Korpipää et al., 2017), and also that comorbid difficulties are common (Amland et al., 2021; Willcutt et al., 2013). This overlap between skills begs the question of what causes such an association. Where some researchers ask whether skills in one domain exert an effect on skills in the other (e.g. Rinne et al., 2020), most research has focused on common underlying abilities (e.g. Koponen et al., 2020; Korpipää et al., 2017). RAN has been utilised as a predictor in several studies on the reading-arithmetic relationship, primarily as it is thought to index common processes related to fluent performance (Koponen et al., 2020).

Understanding how RAN predicts the fluency of reading and math could lead to insights on causes of the apparent overlap between these domains. The multicomponent nature of RAN makes its prediction on either domain of fluency particularly interesting. By using a set of control variables, we can more accurately describe which components of RAN predict which outcomes. In the following, I describe some of the studies that have already attempted to explain how RAN is related to the overlap of reading and math.

7.2 Does the prediction of RAN change over time?

In a cross-sectional study with 216 Israeli primary-schoolers, Balhinez & Shaul (2019) aimed to unravel the extent of the relationships between reading fluency and arithmetic fluency across first-, second- and third grade. They also wanted to find which of the cognitive predictors RAN, inhibition and working memory, would predict performance in the different grades. Overall, they found a moderate association between reading and arithmetic across all the grades. They found that RAN was the most consistent predictor, being a significant predictor in all three grades, and explaining the most unique variance in each type of fluency across grades. The only exception was first-grade reading fluency, which was not predicted by RAN. They suggested that the fluencies are associated, due to similar use of a mechanism for storage and retrieval of associations between verbal-visual stimuli, and that RAN taps retrieval of these representations from memory.

These findings suggest a gradual shift in which underlying abilities are involved – and to what extent – in fluent performance at different developmental stages. In particular, RAN's prediction increases across these stages. It is worth noting however, that the cross-sectional design utilised in this study, means the results across grades reflect three different samples. While it is likely that the results at least partly reflect general developmental trends, characteristics of the samples may affect the results. Further, there were also relatively high rates of attrition in grades 1 and 3, which may have affected the results.

7.3 Is RAN a unique predictor

Whether or not the prediction of RAN is unique, is an important question. If RAN explains variance in reading fluency which is not also explained by other predictors, such as phonological awareness, then RAN explains unique variance. This unique variance suggests that RAN and reading fluency have something in common, which is not shared with phonological awareness. This provides a lot of value in terms of testing whether theoretical explanations are compatible with observed data.

Koponen et al. (2016) found in a longitudinal study (N = 378) that RAN predicted unique variance in later arithmetic fluency and in reading fluency, even after controlling for counting skills, phonological awareness, socio-economic status, verbal short-term memory, working memory and vocabulary. Their interpretation of these results focused on RAN and counting skills as domain-general predictors of fluency. According to them, the fluency of both arithmetic and reading involve similar demands for serial-processing in early stages of development, as well as increasing reliance retrieval of phonological representations. These demands are similar to those involved in counting and RAN, which explain their prediction on fluency.

In contrast, Georgiou et al. (2013) found that whereas RAN accounts for unique variance in reading after controlling for processing speed, phonological awareness and phonological short-term memory, the same was not true for arithmetic fluency. Much of the variance explained by RAN was common between RAN and processing speed. Which led the authors to conclude that the relationship between RAN and arithmetic fluency can be explained by similar demands for processing speed. The contrasting findings between these studies, could for instance be due to Koponen et al. (2017) not controlling for processing speed, or due to differences in the age of the children.

While both studies were longitudinal, Koponen et al. (2017) administered outcome measures in third-grade whereas Georgiou et al. (2013) did so in first-grade. This could mean that the strategies used in arithmetic vary between the samples due to age, as fact retrieval is likely to be more heavily utilised among older children. Thus, the findings of Georgiou et al. (2013) could indicate that RAN's prediction of arithmetic fluency is dependent on developmental stage. Also, RAN could be a unique predictor of fluent arithmetic that is primarily based on fact retrieval, not procedural strategies.

7.4 The role of time and developmental stages

Improvements and shifts in strategy clearly play a part in reading and arithmetic performance. This results in problems when trying to reconcile different findings. Even children in the same grade may utilise qualitatively different strategies with varying degrees of efficiency.

Korpipää et al. (2017) investigated how much of the covariation of reading- and arithmetic fluency was time-specific or time-invariant, and which antecedents would predict each partition of covariance. Children (N = 1335) completed a range of tasks including RAN, working memory and phonological awareness in kindergarten. Then, in first and seventh grade they were measured on timed tasks of reading and arithmetic. The question of time-invariant versus time-specific covariance, is an attempt at understanding how the apparent overlap between reading and arithmetic behaves over time. That is, how much of the overlap is stable and how much of it seems to be restricted to the time of measurement. Korpipää et al. (2017) found that most of the covariance was time-invariant, meaning it is stable over time and not tied to specific ages or developmental stages.

In essence, fluency of arithmetic and reading seem to have common causes that persist over time, but also common causes for why they overlap only in certain stages. If we try to predict these causes from antecedents, then we see that prediction is different for the stable causes and the time-specific ones. The time-invariant portion of the variance was predicted by several predictors including working memory, counting skills and RAN. In first grade the time-specific covariance was predicted by phonological awareness, counting and letter knowledge. While in seventh grade the time-specific covariance was predicted by parental level of education and nonverbal reasoning. This suggests that early - and later fluency - draw upon both common, and specific skills. Some children might for instance perform poorly in both reading and arithmetic through first grade due to poor literacy-related skills, such as letter knowledge or phonological awareness. As these abilities are not predictive of seventh grade covariation, the same children are unlikely to struggle later. However, children who have poor working memory or poor RAN performance are more likely not only to perform poorly in first grade, but also in seventh grade, as these predictors are related to the time-invariant overlap of these skills.

7.5 Does prediction depend on format of RAN?

Hornung et al. (2017) studied the prediction of different formats of RAN on both arithmetic and reading fluency. They administered a battery of seven different RAN tasks at the start of first grade. The stimuli used in these tasks were vowels, consonants, digits, finger-numeral configurations, dice, objects and colours. They constructed different structural equation models where they tested whether a single latent factor, or several that distinguished between specific domains of stimuli, provided a better fit. As all the models fit the data very well, they chose the universal model with one universal factor. Even though other the models also fit, the simpler model is preferable for reasons of parsimony. They also determined that different formats of RAN can lead to better prediction of reading and arithmetic. Alphanumeric RAN explained the most variance in later reading outcomes, whereas two out of three number-specific RAN tasks, namely digits and finger-numeral configurations, but not dice, explained the most variance in arithmetic.

7.6 Explaining the three-way-relation

To understand why RAN seems to predict reading- and arithmetic fluency one must know what underlying processes are common between all three of them. In the following I describe some proposed explanations that detail why RAN predicts the overlap of fluency.

7.6.1 Similarities in strategy development - from procedure to retrieval

The relationship between reading and arithmetic is thought to be at least partly due to a similar pattern of development, where early procedural strategies provide the basis for later use of retrieval (Fuchs et al., 2016). Koponen et al. (2020) describe these early procedural strategies of reading as “serial phonemic assembly”, and their counterparts in arithmetic as “serial reciting of number words”. For reading, this means that phonological recoding provides the basis for “assembly” through unitisation, which in turn is required for retrieval. For arithmetic, the serial reciting of number words is required to employ counting strategies, and to read out problems. The successful implementation of these procedural strategies is key to strengthening problem-solution associations and enabling retrieval.

This means that the fluency of both domains relies similarly on successful serial retrieval - either through grapho-phonemic connections - or through the retrieval of a counting sequence, for consolidation of larger units to take place. The end result being that whole words and sublexical units can be processed and retrieved in reading, and that facts can be retrieved in arithmetic (Koponen et al., 2016). The abilities required for successful implementation of procedural strategies in reading and arithmetic, are here thought to be the same required for serial naming in RAN.

7.6.2 Similar reliance on a mechanism of visual-verbal associations.

Another explanation of why RAN predicts fluency of both reading and arithmetic focuses on the formation of visual-verbal associations in long-term memory. This is not an alternate explanation, but one that focuses on the basis of strategies rather than on the strategies themselves. Koponen et al. (2007) found that RAN, counting and knowledge of letters, predicted the shared variance of reading- and arithmetic fluency. Because these predictors all index visual-verbal associations, they hypothesised that the link between reading and arithmetic concerns the mechanism used to store and retrieve visual-verbal associations. Consequently, the ability to form such links between visual and verbal stimuli is considered central to the relationship between the two domains and is indexed by RAN.

7.6.3 Similar reliance on retrieval of phonological representations

As already outlined, the fluency of both reading and arithmetic could depend on phonological processing. RAN might predict the overlap of fluency between these two domains due to indexing phonological processing. The emphasis here is on RAN indexing the rate of access to phonological representations (Torgesen et al., 1997), or more generally; lexical access (De Smedt, 2018). The implication is nevertheless that RAN, reading fluency and arithmetic fluency might overlap due to similar demands of retrieval of phonological forms (Koponen et al., 2007). This suggests that the overlap of reading fluency and arithmetic fluency is limited to retrieval of phonological forms, primarily of words and of facts. In this case, the prediction of RAN could be mainly due to phonological processing.

7.6.4 Cascaded processing

Koponen et al. (2020) also suggest that RAN, reading- and arithmetic fluency, overlap in demands for cascaded processing. In reading cascaded processing is thought to occur at both a sublexical and a lexical level, depending on whether units are consolidated (Altani et al., 2020), and RAN performance indexes this ability (Protopapas et al., 2013). Koponen et al. (2020) do not elaborate beyond the possibility that serial processing in arithmetic might also involve cascaded processing, and do not explain at which level of arithmetic this would take place.

8 Present study

The present study aims at understanding how rapid automatized naming (RAN) predicts reading fluency and arithmetic fluency, as well as the covariance of fluency. The covariance essentially describes the relationship or the overlap between the domains of fluency, what these domains have in common. RAN is well-established as a unique predictor of reading fluency and has been found to predict arithmetic fluency as well. Using third-grade outcomes is of particular interest as this is likely a period where performance in simple arithmetic will reflect more fluent, fact retrieval-based strategies. Given the relatively transparent nature of the Norwegian orthography, reading accuracy has likely reached a level where it makes sense to capture variation in terms of fluency.

Predicting fluency outcomes with RAN-performance in kindergarten, could yield valuable insights into what underlies the development of fluency in both domains, and help illuminate the overlap between these skills. Further, by controlling for variables such as working memory and processing speed, we can determine how much of RAN's prediction on later outcomes are due to different components of RAN. If RAN predicts unique variance in reading, after controlling for other variables, then there is something in RAN performance which predicts reading beyond shared requirements for working memory and speeded processing. Similarly, the nature of prediction on the overlap between reading fluency and arithmetic fluency; their covariance, could shed light on which underlying processes cause a relationship between the two skills.

Two research questions are at the heart of the present study, where the first leads in to the second:

> *Does RAN predict either reading fluency or arithmetic fluency, separately, after controlling for working memory and processing speed?*
>
> *Does RAN predict the covariance of reading fluency and arithmetic fluency, after controlling for working memory and processing speed?*

9 Methods

The purpose of the current study is to examine the relationship between reading- and arithmetic fluency. Specifically, to see to which extent rapid automatized naming (RAN) predicts fluency after controlling for other variables, namely working memory and processing speed. The study is written in conjunction with the research project: “NumLit: Development of Numeracy and Literacy in Children”, which is a longitudinal research project examining the interplay in development of numeracy and literacy. For present purposes I will utilise data on predictors from kindergarten as well as outcome measures from the third grade.

9.1 Design

The NumLit study is a longitudinal study following the same sample of children from kindergarten until the end of secondary school. The longitudinal design allows the study of changes over time, both between and within participants, on a set of variables (Cohen et al., 2018). Use of the same sample over time enables development to be monitored rather than inferred, which is often done in cross-sectional studies comparing samples of different ages. The children were assessed on a wide range of measures covering skills relating to numeracy, literacy, as well as general cognitive abilities and language. Assessment started in the last year of kindergarten, when children were 5 years old, and has since been repeated every 12 months.

9.2 Sample

The children in the sample were recruited from pre-schools in municipalities around the greater Oslo-area. This area was chosen as it is representative in terms of socioeconomic status, of the national average (Statistics Norway, 2020b, 2020a). Information about the study was provided to parents, and parental consent was given for all participants. 259 children were part of the sample at the first time-point (mean age = 66.1 months, SD = 3.75). At the time of assessment in third grade, the sample consisted of 236 children (mean age = 99.81 months, SD = 3.49). As the study focuses on typical development, children diagnosed with severe learning disabilities or developmental disorders were excluded from the study.

9.3 Data collection

The NumLit study utilises a wide range of tests. During the first three waves of measurement children were assessed on three separate occasions, with three sessions in total across several days. Each assessment lasted between 45-90 minutes.

Due to the outbreak of the Covid-19 pandemic the fourth wave of data collection was less extensive. Children participated in only one session, lasting 60-90 minutes. Permission to commence testing was given by relevant authorities, and a range of preventative measures were employed to hinder risk of infection. Whereas testing at earlier time-points was done by trained research assistants at master's level, this wave of testing was administered by a smaller team, consisting primarily of PhD students from the university of Oslo – specifically the Department of Education and the Department of Special Needs Education.

9.4 Analysis

This study focuses on two theoretical constructs: reading fluency and arithmetic fluency. Theoretical constructs of either form of fluency are not adequately expressed by a single measure. To capture the constructs more holistically, we instead use a range of measures to construct latent factors. In latent variable analysis, measures are treated as the observed indicators of their associated latent constructs (Kline, 2016). The latent factors consist of the covariation between indicators, meaning a latent factor reflects variance that is common between all indicators (Navarro & Foxcroft, 2018).

In this study there are four latent variables of interest: Reading fluency, arithmetic fluency, their covariation, as well as a latent RAN factor. Using structural equation modelling (SEM) we can model relationships between observed and latent variables. Further, we can specify relationships and estimate parameters, such as the effects of RAN on the covariance of fluency, while controlling for working memory and processing speed. The specific type of SEM we will use, is called a structural regression model. In a structural regression model, there is a measurement part and a structural part. The measurement part of the model outlines the relationships of each latent variables with their indicators, while the structural part models the relationships between all variables (Kline, 2016).

Through a series of steps, both the research questions will be answered through structural equation modelling. The measurement part of the model will essentially be the same for both questions, but in order to answer the second question, a second-order factor will also be added to the measurement model. The structural part of the model will be specified to reflect each of the research questions and modified in a justifiable way based on indices of global and local model fit. All the statistical analyses were carried out in R-studio (RStudio Team, 2021) with the lavaan package for latent variable analysis (Rosseel, 2018). Descriptive statistics and graphs were produced in jamovi (The jamovi project, 2021).

9.5 Variables and measurement tools

In this study we are interested in modelling the relationships between pre-school predictors and third-grade outcomes. As such, a range of measures have been employed that will either be used as directly observed variables or as indicators of latent factors. The models will use pre-school measurements as predictors and third-grade measurements as outcomes. In SEM, a distinction between exogenous and endogenous variables is also used. In essence, this describes whether a variable is specified to have a cause within the model, making it endogenous, or if it is without specified causes, making it exogenous (Kline, 2016).

Table 1: List of Measures

Construct	**Measure**	**Test Battery**
Reading Fluency	TOWRE words	TOWRE
	TOWRE pseudo-words	TOWRE
	Oral reading fluency	
Arithmetic Fluency	TOBANS addition	TOBANS
	TOBANS subtraction	TOBANS
	TOBANS addition with carry	TOBANS
	TOBANS subtraction with carry	TOBANS
Rapid automatized naming (RAN)	RAN colours	
	RAN objects	
Working memory	Backward Digit Recall	WMTB-C
Processing speed	Cross-out task	

Note. TOWRE = Test of Word Reading Efficiency, TOBANS = Test of Basic Arithmetic and Numeracy Skills, WMTB-C = Working Memory Test Battery for Children.

9.6 Measures

9.6.1 Rapid automatized naming (pre-school)

Children were tested on two different non-alphanumeric RAN tasks; RAN objects and RAN colours. Non-alphanumeric RAN was chosen because the children had not yet begun numeracy- or literacy instruction. Before each test, children were asked to name a set of four items. For the objects task, these were pictures of a door, a boy, a boat and a mouse (dør, gutt, båt, mus in Norwegian). For the colours task the items were green-, red-, blue- and yellow-coloured circles (grønn, rød, blå, gul in Norwegian). Once children had correctly named each practice item they were presented with the same items in an array, where each of the 4 rows had 8 items. Presentation was pseudorandom such that each row included two configurations of the four items, and configurations would not end with the same item that started the next. Scores were the total time used to name the entire array of items.

9.6.2 Processing speed (pre-school)

The items on the test were pictures of a sun, a boat, a mouse, a door and a bus. The child was given a marker and asked to cross out any mice that they saw. First, the child practiced this once, with a row of seven items. After this they were shown several 6x7 arrays where the target item and distractors items were presented randomly. On the left side, separate from the array, there was a reference picture of the target item. Children were given 60 seconds to cross out as many of the target item as they could across four arrays. After this, a second version of the task was presented with alternate items; a candle, ball, boy, house, and a car. The task was to cross out as many cars as possible in 60 seconds. The total score was the total amount of correctly-crossed-out items over the two tasks.

9.6.3 Working memory (pre-school)

The backward-digit recall task from the working memory test battery for children (WMTB-C) was used to measure working memory capacity. For this task, the test administrator reads aloud a string of digits, which children are asked to repeat backwards. If the presented string was “7, 3” then the correct response would be “3, 7”. The digits were presented orally at intervals of approximately one second. Children must then hold the numbers in their memory and manipulate them in order to generate a correct response. Two practice items proceeded the test. There is a gradual increase in difficulty, as each block of the test consists of six items and each proceeding block adds another digit to the strings. The test was stopped after three mistakes were made in one block. The total score was the total number of correct answers.

9.6.4 TOWRE word-reading fluency (third grade)

Children were told that they were going to read lists of words aloud, and that they should try to read these as quickly and accurately as possible. Further, that they were to read the words on a column-by-column basis. If they reached a word they found difficult, they were allowed to skip the word (which would then be marked as an error). Prior to administering the test, children practiced reading 8 words presented in a column. Once the test started, children were given 45 seconds to read out as many words as possible. Children completed two lists, each containing 26 words of increasing difficulty. Their final score was the average of correctly read words across the two lists.

9.6.5 TOWRE pseudo-word fluency (third grade)

Children were given the same instructions as for word-reading, but this time there was an emphasis on the fact that these “words” were not real words. There were fewer items in the pseudo-word test; 3 columns comprised of 21 pseudo-words of increasing difficulty. Testing and scoring was administered in the exact same way as the word-reading task.

9.6.6 Oral reading fluency (third grade)

Children were told that they were now going to read aloud from two different stories. They should attempt to read as quickly and accurately as possible and continue reading until prompted to stop. The children were told that if they struggled with a word, the tester would read the word for them. This would be done only after they had been stuck on a word for approximately 3 seconds, and the word would then be marked as an error. The title of each story was read aloud to the children before they were given 60 seconds to read from each story. The total amount of correctly read words were noted, and then averaged across the two stories to create the final score.

9.6.7 Measures of arithmetic fluency (third grade)

There were four measures of addition and subtraction fluency. Two of these tested addition and two tested subtraction. The addition tests were administered prior to the subtraction tests to minimise the risk that children would mix the operations.

9.6.8 TOBANS addition and subtraction fluency

Children were told that the goal of this task was to write down answers to math problems as quickly and accurately as possible. They were allowed to skip any problem that they deemed too difficult and could correct any answer if they wished. Prior to the test, children were asked to complete three practice-problems for each type of operation. If they failed to answer at least one of these, they were not asked to complete the measure. After practicing, they were given 60 seconds to complete as many problems as they could. Each test was comprised of 6 total columns, each column containing fifteen single-digit problems that did not involve crossing the ten-boundary. The scores were the total amount of correct answers.

9.6.9 TOBANS addition and subtraction with carry

These tasks matched the previous two in all regards except one. Both the measures were comprised of problems that required crossing the ten-boundary. All the addition problems had sums between 10-18, and in the subtraction task one of the operands was a number between 11-18. Children completed 3 practice problems prior to the task. The task was comprised of 4 total columns, each containing 15 problems. The children were given 60 seconds for each task to write down as many answers as they could.

9.7 Validity

Validity describes the degree to which conclusions, interpretations and measurements are warranted (Cohen et al., 2018). Validity is commonly linked to conclusions and measurements but can be linked to deliberations at all levels of research. A common distinction, outlined by Shadish et al. (2001), is between four types of validity: Internal validity, statistical conclusion validity, construct validity and external validity.

9.7.1 Internal validity

Any conclusion is based on a limited set of observations and a method of tying these together in a coherent manner. The internal validity of a study concerns the degree to which inferences and conclusions regarding the relationships between the variables in question can be sustained (Cohen et al., 2018).

In the present study, our primary aim is to describe the relationship between reading fluency, arithmetic fluency and RAN, these are the primary relationships of interest. Other variables such as working memory and processing speed are employed in order to be able to provide more detailed description of the primary relationships. By using a set of control variables that have been implicated in the literature to possibly explain the prediction of RAN, we can make more accurate, warranted inferences about the primary relationships.

The internal validity of a conclusion describes the extent to which it was reached in a way that makes it defensible in light of rival conclusions (Cohen et al., 2018). For instance, in order to establish the causal effect of one variable on another, several aspects must be considered; correlation, temporality and isolation (Shadish et al., 2001). In a longitudinal, non-experimental study such as NumLit, correlation and temporality can be determined, but not isolation. This means that we can establish whether two variables are related, as well as the directionality of this effect, due to one variable being measured before the other. As the present study is non-experimental, a relationship between two variables could still be caused by unmeasured variables.

In order to rule out other possible explanations, the effect must occur in isolation, meaning modifying the predictor leads to an effect on the outcome (Shadish et al., 2001). In randomised experiments this can be established using control and treatment groups where everything is assumed to be equal, except for the manipulation of the predictor. However, the goal of the current research is not to establish that modifying RAN performance leads to changes in fluency, but to describe whether RAN provides a prediction and how well it is explained by control variables.

9.7.2 External validity

The external validity of a study concerns the degree to which findings are generalisable; whether they also apply in other contexts, are replicable, and the extent to which results from the current sample are likely to reflect a wider population (Cohen et al., 2018). In the current study, several important factors may limit generalisability. For one, if we try to generalise findings to third-graders across all orthographies and educational systems, we fail to consider that reading fluency might develop differently due to qualities of the respective orthography. Differences in educational systems may also mean that instruction differs substantially between countries, or even between classrooms. For instance, Chinese children might be more inclined to use fact retrieval from an early age, as rote learning of arithmetic facts is a common teaching strategy (Georgiou et al., 2020). If two similar studies come to different conclusions, then characteristics of their samples may explain why.

By providing the context and characteristics of a sample, findings might be more valuable to other researchers. It may for instance be easier to compare findings on reading fluency in different orthographies if contextual factors are described. This seems to be common in Finnish research, with researchers stating that the fluency of beginning readers might be influenced by qualities of the orthography (Koponen et al., 2007; Korpipää et al., 2017). Thus, providing context informs the external validity of results, by emphasising the contextual constraints that limit the findings.

9.7.3 Construct validity

Construct validity is all-encompassing. If internal validity is concerned with whether conclusions are based on warranted links between variables, then construct validity asks whether the variables themselves are meaningful (Shadish et al., 2001). For abstract concepts, operationalisation will be through a construct informed by theory (Cohen et al., 2018). The constructs of reading fluency and arithmetic fluency are both theoretical constructs. It is common in quantitative research to make decisions of what measures capture which construct, based on theory, and to look at the convergent and discriminant validity of the indicators to determine whether the choices were warranted (Cohen et al., 2018).

The use of exploratory factor analysis or confirmatory factor analysis, can give insight into whether measures are related, or disassociated to an acceptable extent, given their corresponding theoretical constructs (Navarro & Foxcroft, 2018). Whether or not the indicators load onto the same factor is a common way of determining the convergent and discriminant validity of indicators and latent factors (Kline, 2016).

In the current study we should see that observed measures fit well with their assigned latent factors only. This confirms that the measures display covariance convergent with the assumption that they are all affected by a common latent factor. It does not however confirm that these factors reflect something meaningful, which is a theoretical concern.

9.7.4 Statistical conclusion validity

Links and relations between variables are also subject to the statistical analyses that have been used to find them. Statistical conclusion validity concerns the degree to which relationships are inferred through use of appropriate statistical techniques (García-Pérez, 2012). This is a question of whether the statistical methods that have been employed can logically provide an answer to the research question.

In null-hypothesis significance-testing (NHST) we contrast our research hypothesis with a null-hypothesis of no association between the variables of interest (Navarro & Foxcroft, 2018). Thus, observing a correlation between two variables in our sample does not mean we conclude that the variables are also correlated in the population of interest. Instead, the size of such an effect, the distribution of the data and the size of the sample are used to get an estimate of how distinguishable an effect is from the null-hypothesis; we test for statistical significance (Kline, 2016).

Using appropriate statistical analysis, we get a test statistic describing the observed effect, and an associated p-value which describes how likely our current estimate of the effect is given the null-hypothesis. This is linked to two types of error in NHST; type-I error, the likelihood of incorrectly rejecting a true null-hypothesis; and type-II error, the likelihood of incorrectly retaining a false null-hypothesis (García-Pérez, 2012; Navarro & Foxcroft, 2018).

Type-I error is determined by an alpha-level; an arbitrary threshold of how often we are willing to make a type-I error, most often set to 0.05 (Navarro & Foxcroft, 2018). This means that given the null-hypothesis, any effect size which is deemed to be found in less than 5% of samples is considered significant; unlikely enough given the null-hypothesis that an alternate hypothesis is more likely.

Type-II error is related to the power of the test. The power of a test is affected by the size of the sample, the alpha level, as well as the effect size between the true population parameter and that of the null-hypothesis (Navarro & Foxcroft, 2018). Thus, larger effects can be found to be significant even in small samples, while smaller effect sizes will only be significant in large samples. Both type-I and type-II error are of importance to statistical conclusion validity, as are the assumptions that underly the use of statistical tests (García-Pérez, 2012).

In structural equation modelling (SEM) there is less emphasis on NHST and associated p-values, and an added emphasis on whether a model appreciably fits with the data (Kline, 2016). Models are constructed based on theory and then tested with data in order to see whether the data supports the model. However, a model fitting well with the data does not exclude alternate models, and results from SEM do not lead to a model being confirmed (Kline, 2016). Still, this makes SEM a useful tool for testing hypothetical models based on theory.

The main strength of SEM is the potential to analyse latent factors (Kline, 2016). Where overall least squares regression assumes that measurement is without error (García-Pérez, 2012), in structural regression models each observed endogenous variable is associated with an error term, and each latent factor can also account for unmeasured causes (Kline, 2016). However, one drawback of SEM is that power is generally lower, and large samples may be needed in order to be able to distinguish the fit of a model (Kline, 2016).

9.8 Ethical concerns

The principles of non-maleficence, informed consent and privacy are important principles in educational research (Cohen et al., 2018). The NumLit study has been approved by the Norwegian Centre for Research Data (NSD) as concerns of privacy, ethics and data collection are all addressed in the research design. When children are the subject of research, certain additional concerns should be made. For one, informed consent is contingent on information being available in a form that is comprehensible for children (Cohen et al., 2018). Children in the NumLit study are provided information in an appropriate manner and given the choice to pause or withdraw from testing at any time.

In Norway children under the age of 15 cannot provide consent on their own to participate in research (NSD, n.d.). Parental consent has therefore been collected for all children participating in the NumLit study. Children participate in the study face-to-face and are therefore not anonymous, but all information is strictly confidential. The project does collect some personal information, but no participant should be identifiable from the published data. No information regarding individual participation will be released in any form. As such the project also has rigid rules concerning the storage of data and the protection of personal information. While voice recordings are made during test sessions, this is done only on devices owned by Faculty of Educational Sciences. These recordings are also set to be deleted at a pre-specified date. Further, data is not saved with participant names, but with an ID-code only. In order to link a participant to their ID code, a separate key is needed which is only available on select computers and for specific purposes.

Non-maleficence and beneficence are also important: children should not feel like they are burdened by their participation, rather they should feel that their participation is beneficial (Cohen et al., 2018). In accordance with this, adhering to the testing procedure should not come at the cost of maintaining a working relation with the child during testing, or at the cost of the children’s self-image. Rather participation in the study should be as enjoyable and beneficial as possible.

Due to the Covid-19 pandemic, additional measures were taken to ensure that testing was carried out in a safe manner. If testing put children at elevated risk of infection this would be a serious ethical concern. Parents and schools were therefore contacted and provided information on infection control measures set in place. Testing was carried out by a small team of Master and PhD students, as well as a postdoctoral researcher. The test administrator sat a safe distance from the child, and all materials as well as tables and chairs were disinfected between test sessions.

10 Results

10.1 Descriptive statistics of variables

The third-grade sample consists of 236 children. However, out of these a subset of 219 children provided data for all the variables used in the present study. As maximum likelihood estimation uses only complete data, this means 219 is the effective sample size for structural equation modelling.

Before any statistical analysis is carried out, we attend to the descriptive statistics. The descriptive statistics provide insight into the variables and their distributions. Inspecting descriptive statistics and their associated visual distributions allow us to spot irregularities, determine which indices of central tendency make sense to use, and ensure that the data do not deviate appreciably from the assumptions of the statistical analysis to be employed.

Of particular interest are indices of central tendency, such as the mean or the median which indicate which values the distribution of data is centred around (Navarro & Foxcroft, 2018). Further, we want to describe the dispersion or variability of the data. This can for instance be done in terms of the range or standard deviation (SD); the distance from the mean that corresponds to certain proportions of the data (Navarro & Foxcroft, 2018).

Some descriptive statistics correspond to properties of the distribution of data, namely the skewness and kurtosis. Skewness indicates whether there is an excess of data points on either side of the distribution, or that the distribution trails off to one of the sides (Kline, 2016; Navarro & Foxcroft, 2018). Negative skew indicates that the distribution is “left-tailed”, and that most values in the data are above than the mean, whereas positive skew indicates that the distribution is “right-tailed”, most values being below the mean. Kurtosis describes the “pointedness” of the distribution, where positive kurtosis indicates an excess of data far from the mean, and negative kurtosis indicates a lack of data away from the mean (Navarro & Foxcroft, 2018). Indices of distribution are also often interpreted in conjunction with the Shapiro-Wilk’s statistic W and its associated p-value, which indicates the likelihood that the observed data is normally distributed (Navarro & Foxcroft, 2018).

Where possible, reliability estimates will also be provided. The reliability of a measurement concerns the degree to which it consistently measures what it is intended to measure; to what degree it is free from error (Cohen et al., 2018; Livingston, 2018). One such form of reliability is alternate-forms reliability, which concerns how reliably a participant’s score reflects the score they would have gotten over different versions of the test. Internal consistency reliability is commonly provided, and in many cases can be used as an estimate of the alternate-forms reliability of a measure (Livingston, 2018). Internal consistency reliability concerns how consistently items measure the same thing, the correlation of items with each other (Cohen et al., 2018; Livingston, 2018).

Generally, reliability statistics are reported through the reliability coefficient Cronbach's alpha, or "α", which is a reliability score derived from all possible split-half correlations of items on a test (Navarro & Foxcroft, 2018). McDonald's omega is a more robust internal reliability statistic than α, as it depends on fewer assumptions (Navarro & Foxcroft, 2018). In the following, internal consistency reliability is reported, and for measures where data on individual items are not available, intercorrelations between measures may be provided as a Pearson's r correlation. Both "α" and McDonald's omega "ω" are reported where possible.

Descriptive statistics are reported for the whole sample, rather than just for the subset of 219 analysed in structural equation modelling (SEM). Filtering the variables to only reflect the effective sample does not appreciably change the distributions or the descriptive statistics described below. The N for each variable is reported in table 2.

10.1.1 Transformation of variables

As maximum likelihood estimation assumes both univariate and bivariate normality, and is sensitive to outliers (Kline, 2016), some of the distributions had to be transformed in order to make sure the data adheres to the assumptions of the statistical method. Care was put in to make sure that the transformations did not result in overall lower correlations with other variables, and that they did not adversely affect model fit, as was estimated in confirmatory factor analysis. Further, major disparities between variables in terms of overall variance can lead a model to be underidentified (Kline, 2016). As such some of the transformed variables were multiplied by 10 in order to make sure the SEM would be identified. In the following, where relevant, histograms for untransformed variables are presented on the left, and their transformed counterparts on the right.

10.1.2 RAN objects

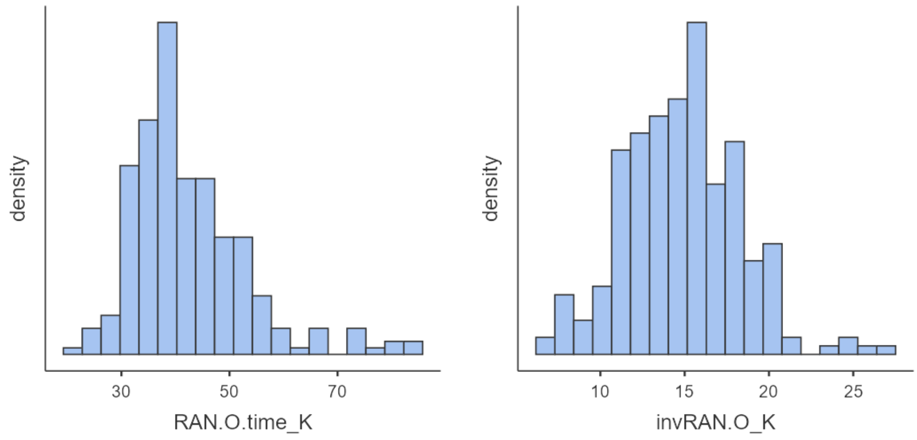


Figure 1: RAN Objects

The distribution of RAN objects is somewhat right-skewed (skewness = 1.37). It has a long tail on the right side as some children took quite a bit longer than the rest to name all the stimuli. Kurtosis is 2.64 indicating that there is an abundance of data points far from the mean. The Shapiro-Wilk test also indicates a significant departure from normality, $W(244) = 0.90$, $p = < .001$.

Due to the skew and kurtosis of the original variable, RAN objects was inversely transformed by dividing 60 by the score. This results in a different metric that reflects how much of the test a child did in 60 seconds. If a child spent 60 seconds, then they achieve a score of 1, and if they spend 75 seconds, they get a score of 0.8. As this transformation greatly decreased the overall variance, the variable it was subsequently multiplied by 10. The transformed variable has a lower skewness of 0.35 and kurtosis of 0.78, the distribution is also more symmetric. The Shapiro-Wilk test still produces a significant p-value, albeit less extreme than for the untransformed variable ($W(244) = 98$, $p = 0.009$). The mean of the variable is 14.92 with a standard deviation of 3.43. There is no available reliability estimate, but the Pearson's correlation between the two transformed RAN measures is $r = .68$.

10.1.3 RAN colours

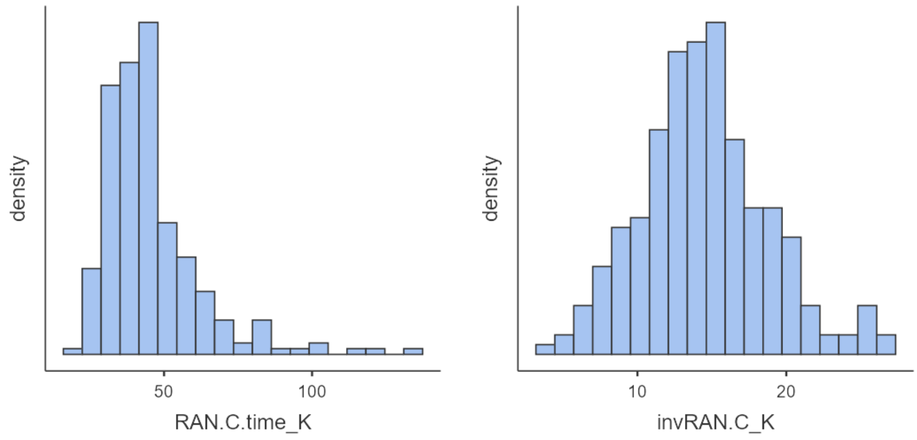


Figure 2: RAN colours

The distribution of RAN colours looks similar to that of RAN objects, but it has an even longer right-trailing tail (skewness = 2.07). It also has an excess of data far from the mean, resulting in a very high kurtosis of 6.20. The Shapiro-Wilk test produces a significant result, $W(244) = 0.83$, $p < .001$. The same inverse transformation that was used for RAN objects was therefore used on RAN colours.

The transformed variable has much lower skewness (0.29), kurtosis (0.21), and produces a non-significant Shapiro Wilk test ($W(244) = 0.99$, $p = 0.067$. The mean of RAN colours is 14.54 (SD = 4.27).

10.1.4 Processing Speed

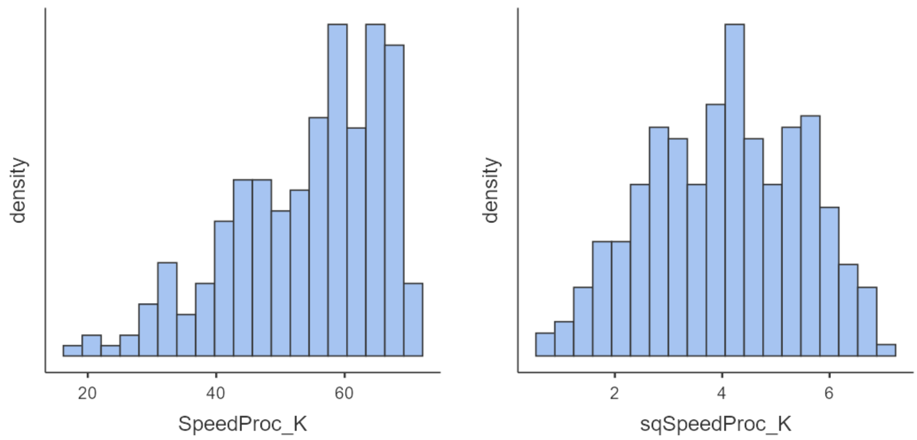


Figure 3: Processing Speed

Processing speed displays an asymmetric distribution with an increasing slope from left to right and a sharp drop at the right side of the distribution. This may indicate that the measure failed somewhat to distinguish between skilled performance. The distribution is somewhat left skewed (skewness = -0.77, kurtosis -0.14). A Shapiro-Wilk test produces a significant result, $W(254) = 0.93$, $p < .001$.

The processing speed variable was transformed using a square-root transformation. This essentially compresses greater scores more than lower ones. The final transformation was "8-sqrt(72-SpeedProc_K)". This means each score starts with a baseline of 8 which is subtracted by the square root of 72 minus the score. This results in a distribution which compresses the right side of the distribution, more than the left side. The new mean is 4.03 (SD = 1.43), skewness is -0.16 and the kurtosis is -0.80. The Shapiro-Wilk test produces a significant but less extreme result ($W(254) = 0.98$, $p = 0.002$). There is no available reliability estimate for this measure.

10.1.5 Backward digit-span

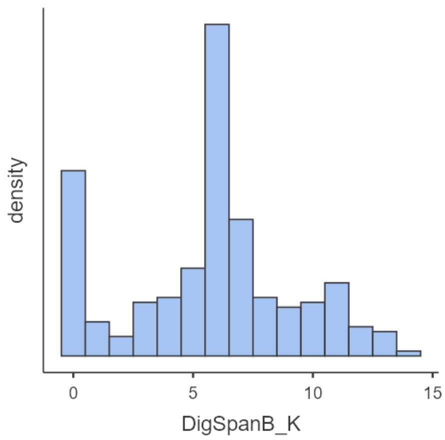


Figure 4: Backward digit-span

The distribution of backward digit-span has a slight floor-effect, giving it a distinct second peak to the far left, meaning several children failed to get even a single item correct. This slight floor effect may result from the instructions of the task being somewhat difficult to grasp for pre-school children. The mean is 5.70 (SD = 3.49), skewness is -0.05 and kurtosis is -0.50. A Shapiro-Wilk test produces a significant result, $W(246) = 0.94$, $p < .001$. Transformation of the variable is not attempted due to the floor effect. The internal reliability estimates for backward digit-span are $\alpha = .89$, and $\omega = .88$.

10.1.6 Test of word reading fluency - words

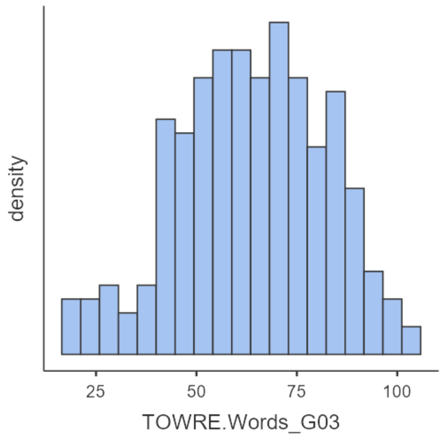


Figure 5: TOWRE words

The distribution of TOWRE words is quite symmetrical, with some slight differences between the far ends of the distribution. Overall, the data looks to fit quite well with the normal distribution. The mean is 63.6 (SD = 18.4), skewness is -0.2,5 and kurtosis is -0.41. A Shapiro-Wilk test comes up significant, W(240) = 0.99, p = 0.031, yet in context of the visual distribution does not seem to indicate a sizeable departure from normality. There is no reliability estimate available for the measure, but the intercorrelation of the two measures used to create this composite is r = .93.

10.1.7 Test of word reading fluency - pseudo-words

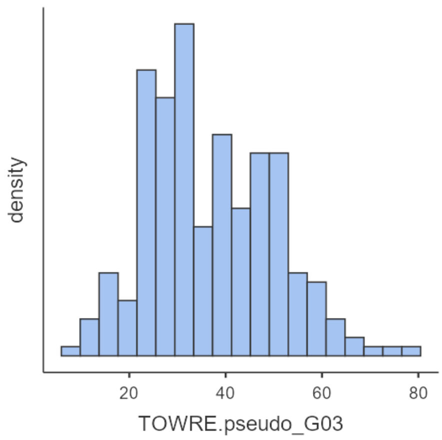


Figure 6: TOWRE pseudo-words

The pseudo-word portion of TOWRE resulted in a variable with a relatively symmetric distribution. There is a distinct peak on the left side of the distribution, and the distribution trails off a bit to the right. The mean is 36.9 (SD = 13.0), skewness is 0.43 and kurtosis is -

0.21. The Shapiro-Wilk test produces a significant result, $W(239) = 0.98$, $p < .001$. The visual distribution and indices of dispersion do not seem to indicate a large deviation from normality. The intercorrelation of the two measures that the variable is comprised of is $r = .93$.

10.1.8 Oral reading fluency

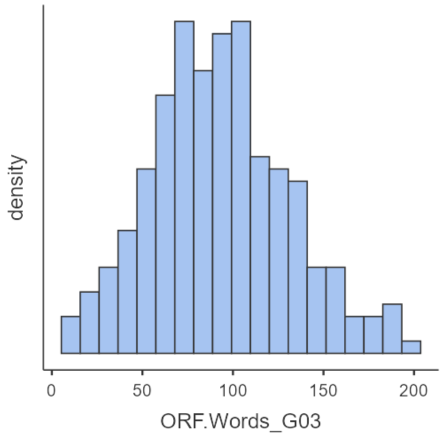


Figure 7: Oral reading fluency

The oral reading fluency measure yielded a distribution that is quite symmetric and approximates a normal distribution. The mean is 92.8 (SD = 38.3), skewness is 0.31 and kurtosis is -0.08. The Shapiro-Wilk test produces a non-significant result, $W(234) = 0.99$, $p = .091$. The intercorrelation of the two oral reading fluency measures used to create this composite variable is $r = .95$.

10.1.9 TOBANS

The TOBANS variables were all transformed with a square-root transformation. However, due to quite a few zero-scores this initially led to lower correlations with some other variables. As a result, before applying the square-root transformation, 2 was added to each score. The scores were also multiplied by 10 to avoid variance discrepancies.

10.1.10 TOBANS addition

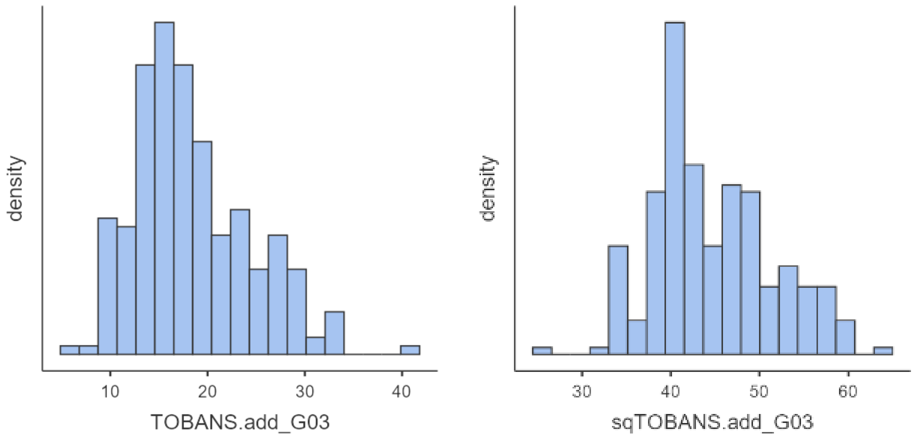


Figure 8: TOBANS addition

The distribution of the TOBANS addition measure is somewhat asymmetric. Only a minor portion of the data is located to the left of the peak, and the distribution is slightly right skewed (skewness = 0.739), with an apparent outlier on the right side (kurtosis = 0.137). The Shapiro-Wilk test produces a significant result, $W(238) = 0.98$, $p < .001$.

The distribution of the transformed TOBANS addition variable looks slightly more symmetric and has lower skew and kurtosis (mean = 44.74, SD = 6.65, N, skewness = 0.40, kurtosis = -0.24). The test still produces a significant, yet less extreme Shapiro-Wilk statistic, $W(238) = 0.98$, $p < .001$. The internal consistency of the test is high with $\alpha = .93$, and $\omega = .93$.

10.1.11 TOBANS addition with carry – 3rd grade

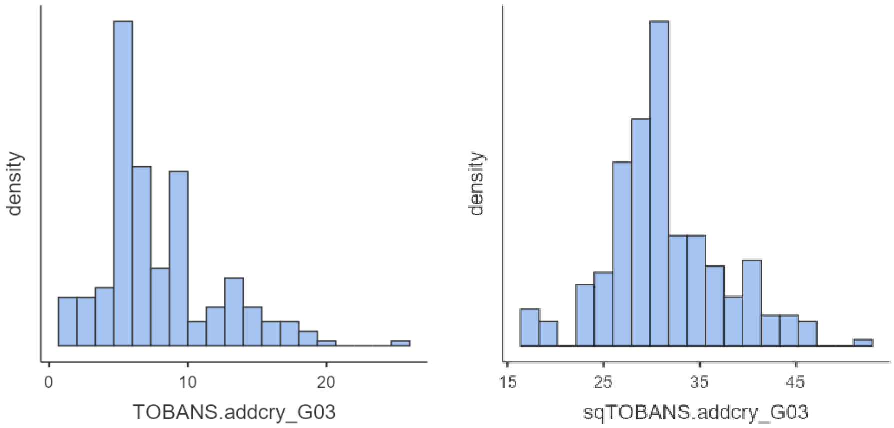


Figure 9: TOBANS addition with carry

The distribution of TOBANS addition with carry is somewhat flat with a few prominent spires. Skewness is 1.03 and kurtosis is 1.16. The Shapiro-Wilk test is significant, $W(238) = 0.93$, $p < .001$.

The distribution of addition with carry after being transformed is more in line with the normal distribution, it also has lower skew and kurtosis (mean = 31.13, SD = 6.26, skewness = 0.44, kurtosis = 0.28). For the transformed variable the Shapiro-Wilk test is still significant, $W(238) = 0.97$, $p < .001$. The internal consistency reliability of the measure is high, with $\alpha = .93$, and $\omega = .92$.

10.1.12 TOBANS subtraction

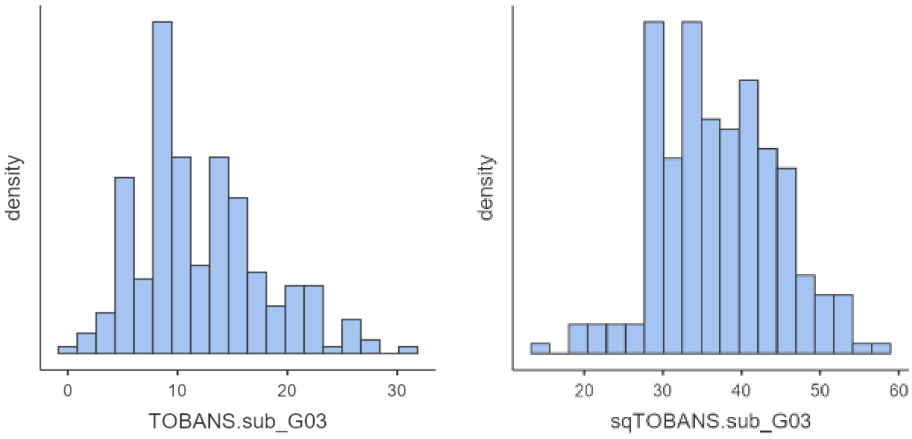


Figure 10: TOBANS subtraction

This distribution has a few gaps between high-points and seems to trail off a bit to the right. Skewness is 0.685 and kurtosis is 0.180. A Shapiro-Wilk test produces a significant result, $W(238) = 0.98$, $p < .001$.

The transformed distribution of subtraction is more symmetric and does not trail similarly to the right (mean = 36.99, SD = 7.45, skewness = 0.16, kurtosis = -0.09). The Shapiro-Wilk test also indicates less of a departure from normality, $W(238) = 0.99$, $p = 0.035$. The internal consistency reliability of the measure is high, $\alpha = 0.88$, and $\omega = .87$.

10.1.13 TOBANS subtraction with carry

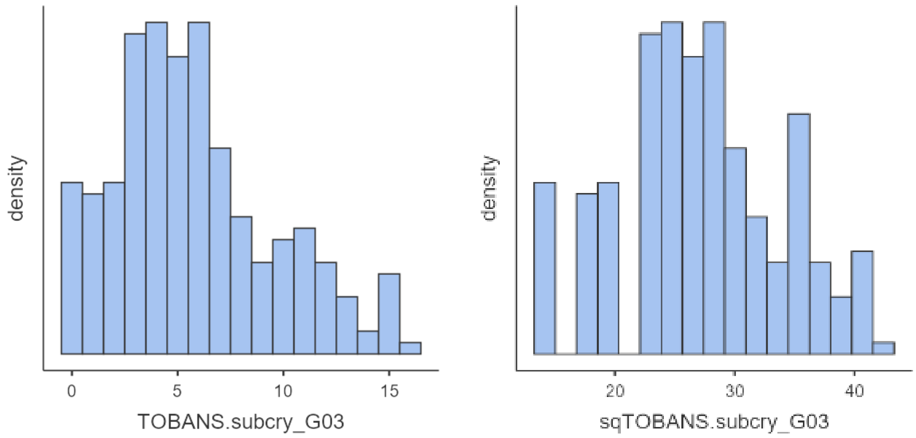


Figure 11: TOBANS subtraction with carry

The subtraction with carry portion of the TOBANS, results in this somewhat asymmetric distribution. Noticeably there are quite a few children who score 0 on this measure, which leads to a slight floor effect. The distribution is quite a bit more compact than the other TOBANS measures, with scores ranging only between 0 and 15. The skewness is 0.648 and kurtosis is -0.154. The Shapiro-Wilk test produces a significant result, W(238) = 0.95, $p < .001$.

The distribution of the transformed variable for subtraction with carry, looks quite similar to the untransformed distribution, has a little lower skew, but a slight increase in kurtosis (mean = 27.02, SD = 6.85, skewness = 0.10, kurtosis -0.49). A Shapiro-Wilk test produces a higher test statistic, but still a low p-value, W(238) = 0.98, $p < .001$. The internal consistency reliability of the measure is α = .93, and ω = .92.

Table 2: Descriptive statistics of the transformed variables

Variable	*N*	*M*	*Mdn*	*SD*	*Range*	*Skewness*	*Kurtosis*	*α*	*ω*
Kindergarten									
Ran Objects	244	14.92	15	3.43	7.06-27.27	0.35	0.78		
Ran colours	244	14.54	14.29	4.27	4.38-27.27	0.29	0.21		
Processing speed	254	4.03	4.13	1.43	0.65-7	-0.16	-0.80		
Backward digit-span	246	5.70	6.00	3.49	0-14	-0.05	-0.50	.89	.88
Third grade									
TOWRE words	240	63.61	64	18.39	18-102.67	-0.25	-0.41		
TOWRE pseudo	239	36.89	34.67	12.97	6.67-77.33	0.43	-0.21		
Oral reading fluency	234	92.84	91.75	38.28	6.50-194.50	0.31	-0.08		
TOBANS addition	238	44.74	43.59	6.65	26.46-64.81	0.40	-0.24	.93	.93
TOBANS subtraction	238	36.29	36.06	7.45	14.14-57.45	0.16	-0.09	.88	.87
TOBANS add carry	238	31.13	30	6.26	17.32-51.96	0.44	0.28	.93	.92
TOBANS sub carry	238	27.02	26.46	6.85	14.14-42.43	0.10	-0.49	.93	.92

Note. N = number of observations, M = mean, Mdn = median, SD = standard deviation, α = Chronbach’s alpha, ω = McDonald’s omega.

10.2 Bivariate correlations

The strength of linear relationships between variables is at the heart of SEM. If we are to model such relationships, then we must first inspect that there are no non-linear relationships, and then gauge the associations between pairs of variables. In other words, we assess the degree of bivariate correlation. In this case we are interested in the degree to which two variables tend towards a linear relationship, which is indexed by Pearson’s r correlation (Navarro & Foxcroft, 2018). One caveat of this index is that a bivariate relationship that is not linear may still yield a high Pearson’s correlation. To ensure that correlations do reflect linear relationships bivariate scatterplots were also assessed, these are also available in the appendix. Spearman’s rho is also provided in the top right section of table 3. This is an index of the degree to which there is an ordinal relationship between the variables (Navarro & Foxcroft, 2018). In the following text, only Pearson’s r correlations are discussed.

The results indicate that there are statistically significant relationships between almost all the variables. Only the processing speed variable displays any non-significant relationships; namely with all the reading measures and with backward digit-span. A non-significant relationship means that the strength of these associations, given the sample size is consistent with the null-hypothesis of no relationship (Navarro & Foxcroft, 2018). All other variables are significantly correlated with p-values of either $< .01$ or $< .001$.

The correlations between measures of reading fluency and arithmetic fluency all range between .33 and .51. The largest of these correlations is that of simple addition and word-reading fluency ($r = .51$, $p < .001$). The effect size of this association even rivals that of simple addition and subtraction with carry ($r = .60$, $p < .001$). In other words, the relationship between addition and subtraction with carry is found to be only slightly greater than that of addition and word-list reading. Clearly there is a sizeable association between the two different domains of fluency.

Both RAN variables display correlations between .40 and .47 with the various reading measures, indicating that there is a moderate-to-strong relationship between RAN performance in kindergarten and reading ability three years later. This is in line with earlier findings on this relationship (e.g. Araújo et al., 2014). The highest among these correlations is that of RAN colours with the oral reading fluency measure ($r = .47$, $p < .001$).

The relationship between RAN and arithmetic seems to be weaker than that of RAN and reading, but is still significant, with correlations ranging from .18 ($p < .01$) to .37 ($p < .001$). Both RAN measures display their highest correlations with simple addition, and lowest correlations with subtraction with carry. The correlations of RAN objects with measures of arithmetic range from .24 to .18, quite a bit lower than those of RAN colours and arithmetic, which range from .37 to .29. These correlations are quite high considering the span of time between measurements. The correlations of the RAN colours variable in particular, are comparable to the overall correlation of .37 between RAN and arithmetic presented in the meta-analysis of Koponen et al. (2017).

Another point of interest is that RAN displays an association with processing speed, .20 ($p < .01$) for RAN objects and .30 ($p <.001$) for RAN colours. However, processing speed shows no significant association with reading fluency, but a weak association with all measures of arithmetic fluency. This begs the question of whether processing speed may explain more of the relationship between RAN and arithmetic than between RAN and reading.

In a similar vein, backward digit-span and the RAN measures are also correlated, $r = .23$ for objects and .29 for colours ($p < .001$ for both). Such relationships can indicate commonalities in terms of processes or causes. However, there is no significant correlation between processing speed and backward digit-span. This indicates that working memory and processing speed, either individually share unmeasured causes with RAN, or are in a causal relationship with RAN, but not with each other.

Interestingly, the correlation between the two forms of non-alphanumeric RAN is .67, ($p < .001$) which is high, but still indicates that performance differences between formats are likely influenced by factors such as familiarity with the stimuli. Higher correlations of RAN colours with both processing speed (.29, $p < .001$), and backward digit-span (.30, $p < .001$), than those for RAN objects (.20, $p < .01$, for processing speed, .23 for backward digit span), may suggest that working memory and processing speed demands are tapped to a somewhat greater extent in colour naming. It would also appear that colour naming more so than object naming is related to arithmetic ability. This latter finding is interesting because it suggests that some features of colour naming correspond more to processes involved in fluent arithmetic.

Table 3: Bivariate correlations between all variables

	RAN objects	RAN colours	Backward digit-span	Processing Speed	TOWRE words	TOWRE pseudo	Oral Reading Fluency	TOBANS Addition	TOBANS addition with carry	TOBANS subtraction	TOBANS subtraction with carry
RAN objects	—	0.65***	0.23***	0.15*	0.39***	0.37***	0.41***	0.26***	0.24***	0.21**	0.19**
RAN colours	0.67***	—	0.30***	0.28***	0.41***	0.40***	0.45***	0.38***	0.32***	0.31***	0.31***
Backward digit-span	0.23***	0.29***	—	0.05	0.23***	0.22***	0.27***	0.20**	0.23***	0.24***	0.18**
Processing speed	0.20**	0.30***	0.06	—	0.10	0.05	0.07	0.22***	0.19**	0.21**	0.23***
TOWRE words	0.41***	0.43***	0.22***	0.09	—	0.88***	0.91***	0.53***	0.46***	0.46***	0.37***
TOWRE pseudo	0.40***	0.40***	0.20**	0.04	0.88***	—	0.85***	0.50***	0.43***	0.48***	0.35***
Oral reading fluency	0.45***	0.47***	0.26***	0.07	0.90***	0.84***	—	0.53***	0.47***	0.49***	0.37***
TOBANS addition	0.24***	0.37***	0.24***	0.23***	0.51***	0.49***	0.52***	—	0.74***	0.78***	0.61***
TOBANS addition with carry	0.21**	0.31***	0.28***	0.17**	0.48***	0.46***	0.48***	0.77***	—	0.73***	0.67***
TOBANS subtraction	0.19**	0.31***	0.27***	0.24***	0.47***	0.49***	0.48***	0.78***	0.75***	—	0.67***
TOBANS subtraction with carry	0.18**	0.29***	0.19**	0.24***	0.37***	0.33***	0.35***	0.60***	0.66***	0.70***	—

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Pearson's r correlations are displayed below the diagonal, Spearman's ρ correlations are displayed above the diagonal. N ranges from 244 to 219 depending on pairwise data availability.

10.3 Structural equation modelling

10.3.1 Indices of global fit

As described in the methods section, structural equation modelling (SEM) will be used to model the relationships between the variables. The type of model that will be used to answer the research questions is a structural regression model. This means first specifying the measurement part of the model and then the structural part. Specifying such a model results in indices of local and global fit. Factor loadings, variances and regression coefficients are indices of local fit, while a range of different indices can be used to determine the global fit of the model (Kline, 2016) Most commonly a set of these are provided together to deliver a comprehensive account of model fit.

In maximum likelihood estimation a fit function is produced based on the discrepancies between covariances predicted by the researcher's model, and the covariances observed in the data (Kline, 2016). Through a series of calculations, the estimation converges on final set of estimates for a model; those that minimise the fit function. The model chi-square (χ^2), which is perhaps the most important of the global fit indices, is calculated with the minimised fit function and sample size. Thus, it compares the fit of the researcher's model with that of a model which perfectly matches the data (Kline, 2016). This χ^2 statistic has an associated p-value, where significant results suggest a poor fit for the model, as this indicates that the researcher's model is inconsistent with the perfectly fitting model.

The root mean square error of approximation (RMSEA), measures department from close fit (Kline, 2016). It is a badness-of-fit statistic where 0 is the best result, which does not indicate that the model fits perfectly, but that it fails to depart from close fit (Kline, 2016). It is typically reported alongside its 90% confidence interval, as these values are also helpful in determining the fit of the model.

The comparative fit index (CFI) compares the amount of departure from close fit for the model in question and a model where all variables are specified to be independent (Kline, 2016). It compares the limit of close fit between the models. A CFI score of .92 means that a model has 92% better fit than the independence model.

Lastly, the standardised root mean square residual (SRMR) is a measure of the average correlation residuals of the model (Kline, 2016). Generally, a number above .10 is seen as indicating poor fit.

10.4 Specifying a structural regression model

10.4.1 The measurement model

I first attended to the measurement part of the SEM. This measurement model will act as a basis for two configurations of structural regression models. First, the latent factors are specified in confirmatory factor analysis (CFA) to determine whether these factors provide appreciable fit with their theoretically assigned factor.

I first specified three latent factors: Reading fluency, with three indicators: TOWRE word-reading, pseudo-word reading, and oral reading fluency. Arithmetic fluency had four indicators: simple addition, simple subtraction, addition with carry and subtraction with carry. The third latent variable was a RAN factor with the indicators of colour RAN and object RAN.

The initial model had acceptable fit ($\chi^2(24) = 35.764$, $p = 0.058$, RMSEA = 0.047, 90% CI [0.000, 0.078], CFI = 0.992, SRMR = 0.028). Inspecting the modification indices, and correlation residuals, there were a few possible changes that would lead to even better fit. However, any such change would need to be theoretically defensible. Changing the model to have correlated residuals between the two subtraction measures, made theoretical sense. This would imply that there is something common between the subtraction measures, that is not effectively captured by arithmetic fluency. For instance, it may be that conceptual understanding of inversion or procedural knowledge relating to specific counting strategies for subtraction are common to these variables. As such elements are not common to all indicators, they are not captured by the arithmetic fluency factor.

The model was therefore modified to include correlated residuals between the subtraction measures, meaning there is room for a relationship between these two variables in the model, even after conditioning them on arithmetic fluency. The resulting model had better fit ($\chi^2(23) = 31.271$, $p = .116$, RMSEA = 0.041, 90% CI [0.000, 0.073], CFI = 0.995, SRMR = 0.026). Next, this measurement model is used as the basis for two structural configurations.

10.4.2 Modelling the first research question – RAN as a predictor of fluency

As the measurement part of the model seems to be correctly specified, the next step is to specify the structural part of the model, corresponding to the first research question. Modelling relationships between variables means turning the CFA model into a structural regression model. First, I specify a path from the RAN factor to each of the fluency factors. After this, I add the observed variables of processing speed and working memory, to the model. These are added to the model along with paths extending to the fluency factors. This specifies that both reading fluency and arithmetic fluency are now regressed on RAN, processing speed and working memory. The resulting model has good fit, $\chi^2(35) = 40.958$, p = .225, RMSEA = 0.028, 90% CI [.000, 0.058], CFI = .996 SRMR =. 0.025.

According to this model RAN is a unique predictor of both reading fluency ($\hat{\beta}$ = 0.569, p < .001) and arithmetic fluency ($\hat{\beta}$ = 0.324, p < .001), after controlling for the effects of working memory and processing speed. Working memory is a significant predictor of arithmetic fluency only ($\hat{\beta}$ = 0.200, p = .004). Processing speed is a non-significant predictor of both forms of fluency, but borders on significant prediction of reading fluency ($\hat{\beta}$ = -0.119, p = .070). Together, the predictors explain 32% of the variance in reading fluency, and 22.5% of the variance in arithmetic fluency. This estimate of explained variance is produced through the coefficient of determination "R^2" (Navarro & Foxcroft, 2018), which in this case is provided by lavaan (Rosseel, 2018).

10.4.3 Modelling the second research question – RAN as a predictor of overlapping fluency

Specifying this model requires taking a step back and testing a different structural specification than the one outlined in the previous section. The starting point is the same measurement model as before, but with an added second-order covariance factor. A second-order factor is one which has no indicators of its own, but is measured indirectly through the indicators of first-order factors, and is specified as a cause of the shared variance between its related first-order factors (Kline, 2016). The second-order factor implies that any association between the two domains of fluency is due to common causes. The modelled covariance factor explains approximately 77% of the variance in reading fluency, and 44% of variance in arithmetic fluency.

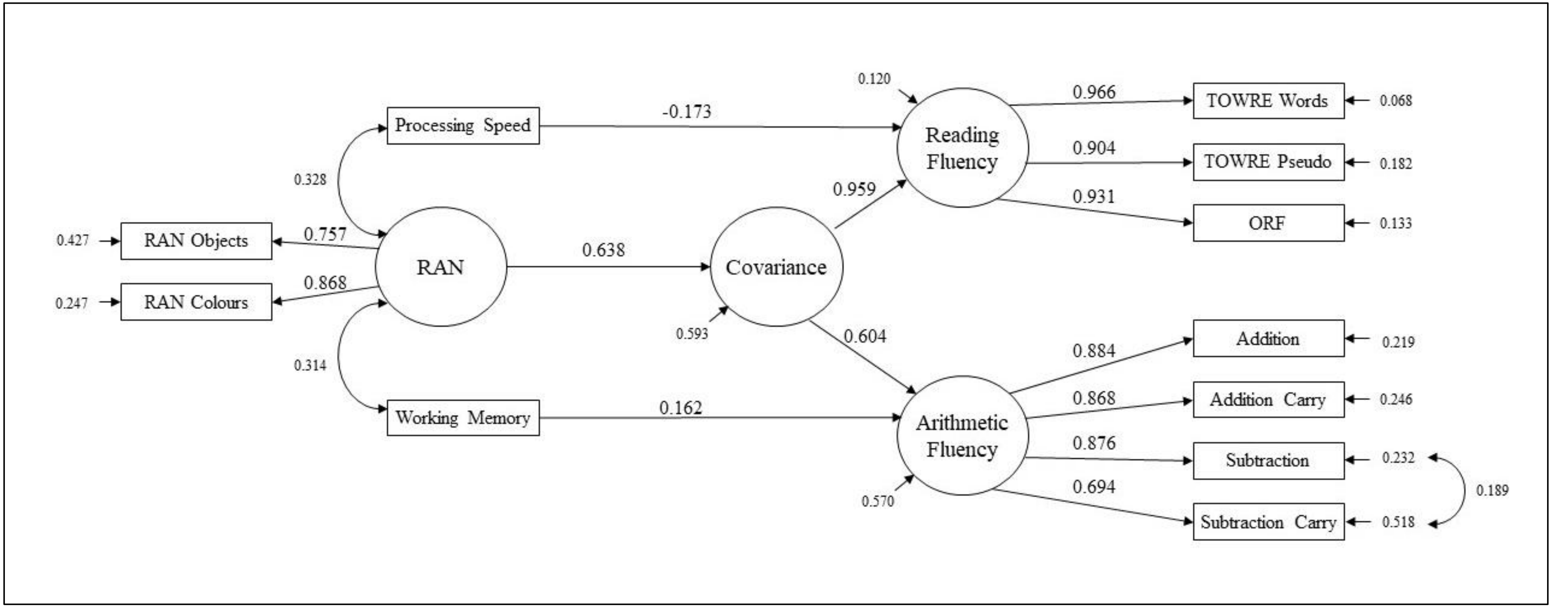
Next, I determine whether the second-order factor is predicted by RAN. I specify that the effect of RAN is on the second-order factor only. This means any effect of RAN on the first-order factors of reading fluency or arithmetic fluency is indirect, and that the second-order factor is regressed on RAN. This step does not adversely affect model fit, and it also produces a significant prediction of RAN ($\hat{\beta}$ = 0.638, p < .001), with RAN explaining 41% of variance in the covariance factor.

After this I test how the control variables fare when predicting the second-order factor. One-by-one I add these variables and specify their effects to be through the covariance factor. Doing this results in a worsened model fit both for processing speed ($\chi_D(1) = 9.33$, $p = .002$), and for working memory ($\chi_D(1) = 5.92$, $p = .015$), as indicated by a chi-square difference test. This suggests that whatever common cause of fluency is predicted by RAN, is not predicted by working memory or processing speed.

Based on this I specify a model where RAN's effect is kept on the second-order covariance factor, but the control variables predict both the first-order factors directly. I then drop the nonsignificant paths of processing speed on arithmetic fluency, of working memory on reading fluency, and the nonsignificant correlation between processing speed and working memory ($r = .048$, $p = .477$). The resulting model (depicted in Figure 12) has good fit, $\chi^2(38) = 44.898$, $p = .205$, RMSEA = 0.029, 90% CI [.000, 0.058], CFI = .996 SRMR = .038.

In order to test whether this model provides worse fit than the model without a second-order factor, that was outlined to answer the first research question, I run a chi-square difference test, which produces a nonsignificant result ($\chi_D(3) = 3.94$, $p = .268$), suggesting that the fit of the models are similar.

Figure 12: Model of overlapping fluency



All estimates are standardised. RAN = Rapid automatized naming, ORF = Oral reading fluency

10.5 Model of fluency

That the model has adequate fit does not mean that it is a correct model, but it means that the model fits the data well. The model with a second-order factor of shared variance between reading fluency and arithmetic fluency (figure 12), presents a hypothesis that the relationship of the two forms of fluency is due to a common latent factor that exerts an effect on both domains of fluency. Further, RAN exerts an effect on this covariance factor. If the data instead suggested that early RAN predicts reading fluency and arithmetic fluency independently, then the model fit would be poorer as there would be residual correlation between RAN and the latent fluency factors.

Overall, the model suggests that a shared causal factor is the reason why there is an association between the fluency of reading and arithmetic. This second-order factor has a very high factor loading on reading fluency ($\hat{\beta} = 0.959$, $p < .001$), and a somewhat lower, but still high loading on arithmetic fluency ($\hat{\beta} = 0.604$, $p < .001$). In other words, whatever causes are common between these factors, seems strongly associated with reading fluency, and somewhat less involved in arithmetic fluency.

The second-order factor is predicted quite well by RAN ($\hat{\beta} = 0.638$, $p < .001$, $R^2 = .41$). This means whatever causes the two domains of fluency to be associated in third grade can be predicted by non-alphanumeric RAN in kindergarten. RAN is here the sole predictor of the second-order factor, meaning the prediction of RAN cannot be explained away by processing speed or working memory.

The RAN factor is significantly correlated with both working memory ($r = 0.31$, $p < .001$) and processing speed ($r = 0.33$, $p < .001$) which in turn are not correlated with each other. This suggests that both processing speed and working memory are implicated in RAN performance, which would make sense seeing as RAN performance involves many different processes.

Each domain of fluency has one direct predictor. For reading fluency this is processing speed ($\hat{\beta} = -0.173$, $p < .01$). And for arithmetic fluency this is working memory ($\hat{\beta} = 0.162$, $p < .01$). These predictions are also interesting. Working memory explains variance in arithmetic fluency, on top of the prediction offered by the second-order factor. This means that working memory provides explanation beyond RAN of what processes are involved in arithmetic fluency. For instance, this could mean that use of derived fact retrieval relies on working memory resources, beyond the working memory aspects involved in RAN.

Somewhat surprisingly, the prediction of the processing speed variable on reading fluency ($\hat{\beta}$ = -0.173, $p < .01$), is negative and significant. As detailed in section 10.2, all the bivariate correlations of processing speed with reading measures, were positive and non-significant. The surprising regression coefficient is due to what is called suppression, which may happen when predictors are correlated. Essentially, this means that the prediction offered by processing speed boosts the prediction of RAN, since the prediction offered by processing speed filters out a part of RAN's prediction that is irrelevant for reading fluency (Friedman & Wall, 2005; Maassen & Bakker, 2001).

The latent factors all have high factor loadings on their indicators. This suggests that whatever is common between the indicators of each latent factor, accounts for a lot of variance in each indicator. Of note, roughly half the variance of the subtraction with carry measure, is explained by the arithmetic fluency factor ($\hat{\beta}$ = 0.694, $p < .001$, $R^2 = .48$), meaning there is quite a bit of unexplained variance, which is also correlated with the unexplained variance of subtraction ($r = .19$, $p < .05$). A similar effect is visible for the RAN factor, where RAN objects performance seems to be affected by unmodeled causes more so than RAN colours.

11 Discussion

11.1 The first research question

Does RAN predict either reading fluency or arithmetic fluency, separately, after controlling for working memory and processing speed?

The findings of the current study largely echo those of previous studies. RAN was found to predict both fluency outcomes, even after controlling for working memory and processing speed. This prediction of RAN is consistent with indirect effects or direct effects on both types of fluency. Before modelling the second order factor, RAN was found to be a unique predictor of both reading fluency ($\hat{\beta}$ = 0.569, p < .001) and arithmetic fluency ($\hat{\beta}$ = 0.324, p < .001) after controlling for both working memory and processing speed. The prediction of the control variables was negligible in comparison to that of RAN. This means that RAN does not seem to predict either type of fluency due to demands for working memory or processing speed. Together, these predictors explained 32% of the variance in reading fluency, and 22.5% of the variance in arithmetic fluency.

11.1.1 Why does RAN predict fluency?

Some possibilities are that RAN predicts the different fluencies due to similar demands for phonological processing, cascaded processing, or orchestration of component processes. Given the findings of previous studies on this relation, it seems unlikely that any single component of RAN performance accounts for the prediction alone, as no study has managed to find a control variable that fully mediates the effect of RAN on reading (Georgiou et al., 2013).

Regarding arithmetic, RAN performance in kindergarten provided a sizeable prediction of arithmetic fluency. This prediction was still smaller than that of RAN on reading fluency. This could indicate that unlike the prediction of reading, fewer components of RAN performance are involved in arithmetic fluency. Phonological processing could be a shared process, implicated through fact retrieval. Cascaded processing somehow seems a less probable candidate in explaining the relation, at least on its own. Solving arithmetic problems does involve some processing of information in sequence, but not to the same extent as reading.

Demands for orchestration could also be involved. While RAN has been described as a condensed version of reading (Norton & Wolf, 2012), there is arguably less overlap of processes involved in RAN and arithmetic. It might still be that the coordination on a smaller scale, of processes such as visual integration and retrieval of phonological representations explains some of the relation of RAN and arithmetic fluency. However, the overall orchestration required for RAN is likely more similar to that of fluent reading, than fluent arithmetic.

11.1.2 Previous research

Working memory and processing speed failed to explain the relation of RAN and either form of fluency. This is in line with earlier research when it comes to reading fluency. However, this does run counter to the findings of Georgiou et al. (2013) who found that RAN's prediction of arithmetic fluency could be explained by demands for processing speed, meaning RAN's prediction was not unique. These different findings may stem from differences in how predictors and outcomes were measured. The Georgiou et al. (2013) study employed two measures of processing speed, where one involved children circling two identical numbers among distractors in an array. Their "calculation fluency" measure required children to make judgements as to whether pairs and trios of digits or configurations of objects summed to a target number. This choice of outcome measure stems from the fact that their sample was in first grade when outcomes were measured. Differences in how processing speed and arithmetic fluency were measured, as well as differing age of participants, could explain the discrepant findings.

11.2 The second research question

Does RAN predict the covariance of reading fluency and arithmetic fluency, after controlling for working memory and processing speed?

The specification of the model in figure 12 reflects a series of steps in which different possible configurations of the model were tested. The model represents both a hypothesis and consequent theoretically defensible modifications. The main hypothesis being tested is that RAN predicts the overlap between reading fluency and arithmetic fluency. The final model suggests that the two forms of fluency do not overlap due to working memory or processing speed ability, but due to reliance on similar processes, that are also involved in RAN.

The observed relationship between reading fluency and arithmetic fluency is modelled to be due to common causes, neither of them exert effects on the other. This means fluent reading does not lead to fluent arithmetic, nor vice versa. Instead they are related through common causes, represented by the covariance factor. The covariance factor does not specify that one single cause is responsible for the association between domains of fluency. The covariance factor simply reflects the relationship between the types of fluency, which can be due to a myriad of common processes. Also, since the predictors were measured in kindergarten, predicting outcomes four years later suggests that at least some aspects of RAN performance are relatively stable over time.

It is hard to pin down which parts of RAN are predictive of reading, of arithmetic or of their overlap. This in turn is why we employ control variables. The idea was to control the prediction of RAN for the effects of both working memory and processing speed. If this appreciably lessened the prediction of RAN, this would inform us that a substantial part of the prediction of RAN would be common with working memory for instance, suggesting that the reason why RAN is predictive is at least partly due to common reliance on working memory.

However, none of the control variables predicted the second-order factor. Instead it seems that processing speed and working memory each directly predict a portion of reading fluency and arithmetic fluency respectively. This means that any processes or combination of processes involved in RAN could be involved in the overlap of fluency. The covariance factor could consist of variance related to phonological processing implicated in both sight-word reading and fact retrieval. This would mean that the phonological processing aspects of RAN would be at the heart of its prediction. In this case, controlling for phonological awareness would reduce the unique prediction of RAN.

However, the high factor loading of the covariance factor on reading fluency, suggests that whatever causes the overlap of fluency also explains most of the variance in reading fluency. As controlling for phonological awareness has repeatedly been found insufficient to explain RAN’s prediction of reading (Georgiou et al., 2013; Landerl et al., 2019; Papadopoulos et al., 2016), it is unlikely that the covariance factor mainly represents phonological processing.

11.2.1 Previous research

The findings of the present study are compatible with those of Koponen et al. (2020), whose model of covariance was the inspiration for the model in figure 12. They found that “serial retrieval fluency", a latent factor indicated by RAN and counting ability RAN predicted shared variance of reading fluency and arithmetic fluency. This factor provided unique prediction of the shared variance of fluency after controlling for a range of predictors including working memory and processing speed. While their serial retrieval factor is different from a RAN factor, the implication of RAN in this factor means that the findings are still compatible.

11.3 Results in a validity context

11.3.1 Internal consistency

The design of the current study does not allow for strong causal inference, as it is a non-experimental study. Causality in this study therefore only indicates the direction of an effect. In SEM, the specification of latent factors assume that they cause the relationship between their indicators (Kline, 2016). This does not necessarily account for all variables that may be involved in the true causal relation. Modelling structural relations, for instance RAN as a "cause" of the covariance of fluency, does not mean we can conclude that improving RAN performance would improve fluent performance in reading and arithmetic. Instead we can only describe the developmental relationship between the variables: It would appear that better performance on non-alphanumeric RAN in kindergarten, leads to more fluent reading and arithmetic in third grade. However, this does not rule out rival explanations, such as unmeasured variables causing the relationship of RAN and fluency.

The complex nature of RAN performance further complicates prediction. Any conclusions regarding what causes the fluency overlap need to address the ambiguity of RAN as a predictor. Control variables provide an avenue of addressing this, and in this study, we can say that RAN's prediction is not explained by working memory or processing speed.

11.3.2 External validity

If this study was replicated with a comparable population, but in a different context, the results might not replicate. There are several factors that limit generalisation. The fluency of reading and arithmetic are both theorised to undergo shifts in which underlying processes are central at different stages of development (Balhinez & Shaul, 2019). This means that the predictors of third-grade fluency may be completely different from earlier or later predictors.

In this regard, the present study represents a developmental snapshot at a certain point of development, and reflects only one of several configurations the association between

RAN, reading fluency and arithmetic fluency, will undergo. Therefore, the findings likely only generalise to third graders, and even for this group, cultural differences could mean development progresses differently.

As detailed in the section on reading fluency, the current study was carried out in a semi-transparent orthography. If a similar study was carried out in a different orthography, the findings may differ, as a consequence of differences in developmental progression relating to qualities of the written language. Norwegian third graders might rely differently on reading strategies than English third graders. English-speaking children also tend to start literacy instruction at an earlier point due to complexity of the written language.

The present study was carried out in a convenience sample, meaning there may be systematic differences between the sample and population it is thought to represent. Kindergartens were selected at random, but children had to opt into the study given parental consent. Selection of participants not being random, but rather being opt-in, could mean that participants might differ from the wider population on unmeasured variables that could be related to the variables of interest, such as socio-economic status. While possibly leading to a slight bias in the results, the method of sampling is consistent with an ethically informed approach, a concern which takes precedence.

11.3.3 Statistical conclusion validity

In order to ensure the statistical conclusion validity of the current study many considerations were made: Variables were transformed to make sure they did not violate the assumptions of SEM. The specification of a SEM was carried out through a series of steps to ensure that the final model had adequate fit. Parameter estimates played a secondary part in this process, being considered only if model fit was satisfactory.

The final model is consistent with both excellent fit, and not-close fit, as the upper interval of the 90% CI interval of the RMSEA is above 0.05 (Kline, 2016). It is, however, not consistent with poor fit as this value is below .10. If the power of the study was greater, the fit of the model might be more readily distinguished, yet the current results do not indicate in any way that model fit is poor.

The prediction of RAN on the covariance of fluency has an associated p-value of less than .001, this means that given the sample size, and if the null-hypothesis was true, there would be less than 0.1% chance of observing such a large effect. However, the p-values of working memory’s effect on arithmetic fluency were larger, meaning there is a slightly greater chance of committing a type-I error. Omitted paths in the model could also potentially reflect type-II error, as the power of the study was too low to detect these effects. However, omitted paths also reflect considerations of model fit.

A possible threat to the statistical validity of the study is that the model presented is just one of several possible models that would support the data equally well or better. Conclusions that claim to inhabit statistical validity must therefore concede that this is simply one of several possible ways to explain the relationship of these variables. The steps taken to arrive on the final model do, however, mean that the model can be preferred over other similar models which proved to have poor fit or nonsignificant parameter estimates.

11.3.4 Construct validity

Arguably one of the most limiting factors in terms of construct validity are the control variables. While backward digit-span is often used in research as a measure of working memory (e.g. Koponen et al., 2016; Korpipää et al., 2017), this may lead to a limited construct of working memory. The same may be true for processing speed. Findings may have been different if more indicators were used for these constructs. Still, it is not always better to use more measurements. If one measurement has good psychometrics, it might be better to rely on just one, rather than to construct latent factors with other indicators that have worse psychometric properties (Kline, 2016).

The results of this study warrant mentioning the fallacies of naming and reification. The naming fallacy corresponds to the assumption that because a latent factor has been given a name, it corresponds to a hypothetical construct, whereas the reification fallacy is tied to the assumption that a factor must represent a real thing (Kline, 2016).

It is probably fair to say that the latent factor of reading fluency can be used to represent its corresponding theoretical construct. The indicators are selected as they all represent different forms of fluent reading. While they all display unique variance tied to specific demands of the task, their shared variance seems a good approximation of reading fluency.

The same is true for arithmetic fluency, although this construct is less readily defined in the literature. While we assume that fact retrieval plays a part in fluent performance, the factor modelled in this study likely corresponds more to general definitions of arithmetic fluency; those that do not see arithmetic fluency as synonymous with frequency of fact retrieval.

In interpreting the covariance factor, we should steer clear of a reification fallacy. This is one of the reasons why no clear name has been given to this factor. It is simply labelled as the covariance of fluency since we cannot clearly delineate what it is. Predicting it does provide insight as to why there is an overlap of fluency. It is also possible that the covariance factor picks up common method variance, meaning that at least some of the shared variance of fluent reading and arithmetic is an artefact of the method of measurement (Kline, 2016). If this was the case, then the processing speed variable should have also predicted the covariance factor, as processing speed, similarly to RAN was a timed measure.

The fit of the measurement part of the model shows that the factors demonstrate convergent and discriminant validity. The latent factors display convergent validity, as there are sizeable correlations between indicators of the same construct. Also, each factor has high factor loadings on their indicators. They also display discriminant validity as the indicators of different factors do not display large intercorrelations. However, factors do not necessarily perfectly mirror their theoretical counterparts.

12 Conclusion

Pre-school rapid automatized naming predicted the fluency of both reading and arithmetic in the third grade. The relationship between fluent reading and arithmetic suggests that they overlap in terms of underlying processes. Through structural equation modelling we find that this overlap is predicted by non-alphanumeric RAN, but not by working memory or processing speed. This suggests that arithmetic fluency and reading fluency do not overlap due to working memory or processing speed demands, but due to processes involved in RAN. Beyond this, working memory seems to be implicated in ability specific to arithmetic fluency. Speed of processing acted as a suppressor, indicating that the shared aspects of RAN and speed of processing, are not the parts of RAN that predict reading fluency.

Overall, this suggests that the development of certain processes involved in RAN are stable, and that these correspond to processes that contribute to fluent reading and arithmetic in third grade.

12.1 Implications for Special Needs Education

The findings of the current study suggest that non-alphanumeric RAN could be valuable as an early screening measure. Specifically, screening for poor RAN performance in kindergarten may help identify children who are at risk of difficulties relating to fluent reading and arithmetic. Children with poor RAN performance might benefit from preventative measures in reading and math instruction. However, as the current sample is a typically developing sample, we cannot know whether the relationships between predictors and outcomes would be different for children with specific learning disabilities.

These findings should not be taken as indication that improving RAN performance would have any merit as a fluency intervention. Further investigation of the relationship between RAN performance and fluency and might pinpoint more accurately common processes, which in turn could be better candidates for intervention studies.

The relationship between fluency of reading and arithmetic, also suggests that when children are identified as having word-level reading difficulties, it might be prudent to evaluate whether they also struggle with arithmetic fluency, and vice versa. Simply evaluating the accuracy of arithmetic performance might lead to underidentification of difficulties relating to arithmetic fluency and fact retrieval.

12.2 Limitations

The present study is conducted in a semi-transparent orthography. Since qualities of orthography might affect reading development, the relations of reading fluency and arithmetic fluency might be different across orthographies. Cultural factors may also be at play, affecting the relationships between the variables. Further, the method of sampling could bias the results somewhat.

As RAN performance involves several subprocesses, we cannot describe exactly why RAN predicts reading fluency and arithmetic fluency. If additional control variables had been used, it may have been possible to give more precise descriptions of these relations. Additional relevant predictors could also elucidate what abilities contribute to fluent performance, beyond those involved in RAN.

13 References

Altani, A., Georgiou, G. K., Deng, C., Cho, J.-R., Katopodi, K., Wei, W., & Protopapas, A. (2017). Is processing of symbols and words influenced by writing system? Evidence from Chinese, Korean, English, and Greek. *Journal of Experimental Child Psychology*, *164*, 117–135. https://doi.org/10.1016/j.jecp.2017.07.006

Altani, A., Protopapas, A., Katopodi, K., & Georgiou, G. K. (2020). From individual word recognition to word list and text reading fluency. *Journal of Educational Psychology*, *112*(1), 22–39. https://doi.org/10.1037/edu0000359

Amland, T., Lervåg, A., & Melby-Lervåg, M. (2021). Comorbidity Between Math and Reading Problems: Is Phonological Processing a Mutual Factor? *Frontiers in Human Neuroscience*, *14*. https://doi.org/10.3389/fnhum.2020.577304

Andres, M., & Pesenti, M. (2015). Finger-based representation of mental arithmetic. *The Oxford Handbook of Numerical Cognition*. https://doi.org/10.1093/oxfordhb/9780199642342.013.028

Araújo, S., & Faísca, L. (2019). A Meta-Analytic Review of Naming-Speed Deficits in Developmental Dyslexia. *Scientific Studies of Reading*, *23*(5), 349–368. https://doi.org/10.1080/10888438.2019.1572758

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2014). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology*, *107*(3), 868. 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*(2), 177–201. https://doi.org/10.1002/rrq.159

Balhinez, R., & Shaul, S. (2019). The Relationship Between Reading Fluency and Arithmetic Fact Fluency and Their Shared Cognitive Skills: A Developmental Perspective. *Frontiers in Psychology*, *10*, 1281. https://doi.org/10.3389/fpsyg.2019.01281

Bowers, P. G., & Wolf, M. (1993). Theoretical links among naming speed, precise timing mechanisms and orthographic skill in dyslexia. *Reading and Writing*, *5*(1), 69–85. https://doi.org/10.1007/BF01026919

Butterworth, B., Varma, S., & Laurillard, D. (2015). Dyscalculia: From brain to education. *The Oxford Handbook of Numerical Cognition*. https://doi.org/10.1093/oxfordhb/9780199642342.013.001

Cain, K. (2010). *Reading development and difficulties*. BPS Blackwell/John Wiley.

Campbell, J. I. D. (1994). Architectures for numerical cognition. *Cognition*, *53*(1), 1–44. https://doi.org/10.1016/0010-0277(94)90075-2

Campbell, J. I. D. (2015). How Abstract is Arithmetic? *The Oxford Handbook of Numerical Cognition*. https://doi.org/10.1093/oxfordhb/9780199642342.013.024

Carr, M., & Alexeev, N. (2011). Fluency, accuracy, and gender predict developmental trajectories of arithmetic strategies. *Journal of Educational Psychology*, *103*(3), 617. https://doi.org/10.1037/a0023864

Chong, S. L., & Siegel, L. S. (2008). Stability of Computational Deficits in Math Learning Disability From Second Through Fifth Grades. *Developmental Neuropsychology*, *33*(3), 300–317. https://doi.org/10.1080/87565640801982387

Cirino, P. T., Child, A. E., & Macdonald, K. T. (2018). Longitudinal predictors of the overlap between reading and math skills. *Contemporary Educational Psychology*, *54*, 99–111. https://doi.org/10.1016/j.cedpsych.2018.06.002

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (Eighth edition). Routledge.

Cowan, R., Donlan, C., Shepherd, D.-L., Cole-Fletcher, R., Saxton, M., & Hurry, J. (2011). Basic calculation proficiency and mathematics achievement in elementary school children. *Journal of Educational Psychology*, *103*(4), 786–803. https://doi.org/10.1037/a0024556

De Smedt, B. (2018). Language and Arithmetic. In *Heterogeneity of Function in Numerical Cognition* (pp. 51–74). Elsevier. https://doi.org/10.1016/B978-0-12-811529-9.00003-0

De Smedt, B., & Boets, B. (2010). Phonological processing and arithmetic fact retrieval: Evidence from developmental dyslexia. *Neuropsychologia*, *48*(14), 3973–3981. https://doi.org/10.1016/j.neuropsychologia.2010.10.018

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1–2), 1–42. https://doi.org/10.1016/0010-0277(92)90049-N

Diamanti, V., Goulandris, N., Campbell, R., & Protopapas, A. (2018). Dyslexia Profiles Across Orthographies Differing in Transparency: An Evaluation of Theoretical Predictions Contrasting English and Greek. *Scientific Studies of Reading*, *22*(1), 55–69. https://doi.org/10.1080/10888438.2017.1338291

Donker, M., Kroesbergen, E., Slot, E., Van Viersen, S., & De Bree, E. (2016). Alphanumeric and non-alphanumeric Rapid Automatized Naming in children with reading and/or spelling difficulties and mathematical difficulties. *Learning and Individual Differences*, *47*, 80–87. https://doi.org/10.1016/j.lindif.2015.12.011

Dowker, A. (2014). Young children's use of derived fact strategies for addition and subtraction. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00924

Dowker, A. (2015). Individual Differences in Arithmetical Abilities. *The Oxford Handbook of Numerical Cognition*. https://doi.org/10.1093/oxfordhb/9780199642342.013.034

Ehri, L. C. (2014). Orthographic Mapping in the Acquisition of Sight Word Reading, Spelling Memory, and Vocabulary Learning. *Scientific Studies of Reading*, *18*(1), 5–21. https://doi.org/10.1080/10888438.2013.819356

Foley, A. E., Vasilyeva, M., & Laski, E. V. (2017). Children's use of decomposition strategies mediates the visuospatial memory and arithmetic accuracy relation. *British Journal of Developmental Psychology*, *35*(2), 303–309. https://doi.org/10.1111/bjdp.12166

Friedman, L., & Wall, M. (2005). Graphical Views of Suppression and Multicollinearity in Multiple Linear Regression. *The American Statistician*, *59*(2), 127–136. https://doi.org/10.1198/000313005X41337

Fuchs, L. S., Geary, D. C., Fuchs, D., Compton, D. L., & Hamlett, C. L. (2016). Pathways to Third-Grade Calculation Versus Word-Reading Competence: Are They More Alike or Different? *Child Development*, *87*(2), 558–567. https://doi.org/10.1111/cdev.12474

García-Pérez, M. A. (2012). Statistical Conclusion Validity: Some Common Threats and Simple Remedies. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00325

Geary, D. C., Hoard, M. K., & Bailey, D. H. (2012). Fact Retrieval Deficits in Low Achieving Children and Children With Mathematical Learning Disability. *Journal of Learning Disabilities*, *45*(4), 291–307. https://doi.org/10.1177/0022219410392046

Georgiou, G. K., & Parrila, R. (2020). What mechanism underlies the rapid automatized naming–reading relation? *Journal of Experimental Child Psychology*, *194*. https://doi.org/10.1016/j.jecp.2020.104840

Georgiou, G. K., Parrila, R., & Kirby, J. R. (2009). RAN Components and Reading Development From Grade 3 to Grade 5: What Underlies Their Relationship? *Scientific Studies of Reading*, *13*(6), 508–534. https://doi.org/10.1080/10888430903034796

Georgiou, G. K., Parrila, R., & Papadopoulos, T. C. (2008). Predictors of word decoding and reading fluency across languages varying in orthographic consistency. *Journal of Educational Psychology*, *100*(3), 566–580. https://doi.org/10.1037/0022-0663.100.3.566

Georgiou, G. K., Tziraki, N., Manolitsis, G., & Fella, A. (2013). Is rapid automatized naming related to reading and mathematics for the same reason(s)? A follow-up study from kindergarten to Grade 1. *Journal of Experimental Child Psychology*, *115*(3), 481–496. https://doi.org/10.1016/j.jecp.2013.01.004

Georgiou, G. K., Wei, W., Inoue, T., & Deng, C. (2020). Are the relations of rapid automatized naming with reading and mathematics accuracy and fluency bidirectional? Evidence from a 5-year longitudinal study with Chinese children. *Journal of Educational Psychology*, *112*(8), 1506–1520. https://doi.org/10.1037/edu0000452

Gilmore, C., Göbel, S. M., & Inglis, M. (2018). *An introduction to mathematical cognition*. Routledge,Taylor & Francis Group.

Hornung, C., Martin, R., & Fayol, M. (2017). General and Specific Contributions of RAN to Reading and Arithmetic Fluency in First Graders: A Longitudinal Latent Variable Approach. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01746

Imbo, I., & Vandierendonck, A. (2007). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology*, *96*(4), 284–309. https://doi.org/10.1016/j.jecp.2006.09.001

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003a). A Longitudinal Study of Mathematical Competencies in Children With Specific Mathematics Difficulties Versus Children With Comorbid Mathematics and Reading Difficulties. *Child Development*, *74*(3), 834–850. https://doi.org/10.1111/1467-8624.00571

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003b). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology*, *85*(2), 103–119. https://doi.org/10.1016/S0022-0965(03)00032-8

Kail, R., & Hall, L. K. (1994). Processing speed, naming speed, and reading. *Developmental Psychology*, *30*(6), 949–954. https://doi.org/10.1037/0012-1649.30.6.949

Kail, R., Hall, L. K., & Caskey, B. J. (1999). Processing speed, exposure to print, and naming speed. *Applied Psycholinguistics*, *20*(2), 303–314. https://doi.org/10.1017/S0142716499002076

Karagiannakis, G., Baccaglini-Frank, A., & Papadatos, Y. (2014). Mathematical learning difficulties subtypes classification. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00057

Kaufmann, L., Lochy, A., Drexler, A., & Semenza, C. (2004). Deficient arithmetic fact retrieval—Storage or access problem?: A case study. *Neuropsychologia*, *42*(4), 482–496. https://doi.org/10.1016/j.neuropsychologia.2003.09.004

Kim, Y.-S. (2015). Developmental, Component-Based Model of Reading Fluency: An Investigation of Predictors of Word-Reading Fluency, Text-Reading Fluency, and Reading Comprehension. *Reading Research Quarterly*, *50*(4), 459–481. https://doi.org/10.1002/rrq.107

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (Fourth edition). The Guilford Press.

Koponen, T., Aunola, K., Ahonen, T., & Nurmi, J.-E. (2007). Cognitive predictors of single-digit and procedural calculation skills and their covariation with reading skill. *Journal of Experimental Child Psychology*, *97*(3), 220–241. https://doi.org/10.1016/j.jecp.2007.03.001

Koponen, T., Eklund, K., Heikkilä, R., Salminen, J., Fuchs, L., Fuchs, D., & Aro, M. (2020). Cognitive Correlates of the Covariance in Reading and Arithmetic Fluency: Importance of Serial Retrieval Fluency. *Child Development*, *91*(4), 1063–1080. https://doi.org/10.1111/cdev.13287

Koponen, T., Georgiou, G., Salmi, P., Leskinen, M., & Aro, M. (2017). A meta-analysis of the relation between RAN and mathematics. *Journal of Educational Psychology*, *109*(7), 977–992. https://doi.org/10.1037/edu0000182

Koponen, T., Salmi, P., Torppa, M., Eklund, K., Aro, T., Aro, M., Poikkeus, A.-M., Lerkkanen, M.-K., & Nurmi, J.-E. (2016). Counting and rapid naming predict the fluency of arithmetic and reading skills. *Contemporary Educational Psychology*, *44–45*, 83–94. https://doi.org/10.1016/j.cedpsych.2016.02.004

Korpipää, H., Koponen, T., Aro, M., Tolvanen, A., Aunola, K., Poikkeus, A.-M., Lerkkanen, M.-K., & Nurmi, J.-E. (2017). Covariation between reading and arithmetic skills from Grade 1 to Grade 7. *Contemporary Educational Psychology*, *51*, 131–140. https://doi.org/10.1016/j.cedpsych.2017.06.005

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*(2), 230–251. https://doi.org/10.1598/RRQ.45.2.4

Landerl, K., Freudenthaler, H. H., Heene, M., De Jong, P. F., Desrochers, A., Manolitsis, G., Parrila, R., & Georgiou, G. K. (2019). Phonological Awareness and Rapid Automatized Naming as Longitudinal Predictors of Reading in Five Alphabetic Orthographies with Varying Degrees of Consistency. *Scientific Studies of Reading*, *23*(3), 220–234. https://doi.org/10.1080/10888438.2018.1510936

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children’s learning of multiplication. *Journal of Experimental Psychology: General*, *124*(1), 83–97. https://doi.org/10.1037/0096-3445.124.1.83

Lervåg, A., & Hulme, C. (2009). Rapid Automatized Naming (RAN) Taps a Mechanism That Places Constraints on the Development of Early Reading Fluency. *Psychological Science*, *20*(8), 1040–1048. https://doi.org/10.1111/j.1467-9280.2009.02405.x

Livingston, S. A. (2018). *Test Reliability—Basic Concepts* (Research Memorandum No. RM-18-01). Princeton NJ: Educational Testing Service.

Maassen, G. H., & Bakker, A. B. (2001). Suppressor Variables in Path Models: Definitions and Interpretations. *Sociological Methods & Research*, *30*(2), 241–270. https://doi.org/10.1177/0049124101030002004

Malone, S. A., Heron-Delaney, M., Burgoyne, K., & Hulme, C. (2019). Learning correspondences between magnitudes, symbols and words: Evidence for a triple code model of arithmetic development. *Cognition*, *187*, 1–9. https://doi.org/10.1016/j.cognition.2018.11.016

Moll, K., Fussenegger, B., Willburger, E., & Landerl, K. (2009). RAN Is Not a Measure of Orthographic Processing. Evidence From the Asymmetric German Orthography. *Scientific Studies of Reading*, *13*(1), 1–25. https://doi.org/10.1080/10888430802631684

Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., Streiftau, S., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., Tóth, D., Honbolygó, F., Csépe, V., Bogliotti, C., Iannuzzi, S., Démonet, J.-F., Longeras, E., Valdois, S., George, F., … Landerl, K. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learning and Instruction*, *29*, 65–77. https://doi.org/10.1016/j.learninstruc.2013.09.003

Navarro, D. J., & Foxcroft, D. R. (2018). *Learning statistics with jamovi: A tutorial for psychology students and other beginners*. Danielle J. Navarro and David R. Foxcroft. https://doi.org/10.24384/HGC3-7P15

Norton, E. S., & Wolf, M. (2012). Rapid Automatized Naming (RAN) and Reading Fluency: Implications for Understanding and Treatment of Reading Disabilities. *Annual Review of Psychology*, *63*(1), 427–452. https://doi.org/10.1146/annurev-psych-120710-100431

NSD. (n.d.). *Barnehage- og skoleforskning*. NSD. Retrieved 30 March 2021, from https://nsd.no/personverntjenester/oppslagsverk-for-personvern-i-forskning/barnehage-og-skoleforskning

OECD & Statistics Canada. (2011). *Literacy for Life: Further Results from the Adult Literacy and Life Skills Survey*. OECD. https://doi.org/10.1787/9789264091269-en

Papadopoulos, T. C., Spanoudis, G. C., & Georgiou, G. K. (2016). How Is RAN Related to Reading Fluency? A Comprehensive Examination of the Prominent Theoretical Accounts. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01217

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*, *21*(3), 235–253. https://doi.org/10.1002/dys.1500

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading*, *11*(4), 357–383. https://doi.org/10.1080/10888430701530730

Polspoel, B., Peters, L., Vandermosten, M., & Smedt, B. D. (2017). Strategy over operation: Neural activation in subtraction and multiplication during fact retrieval and procedural strategy use in children. *Human Brain Mapping*, *38*(9), 4657–4670. https://doi.org/10.1002/hbm.23691

Priya, K., & Wagner, R. K. (2009). Development of reading comprehension. In R. K. Wagner, C. Schatschneider, & C. Phythian-Sence (Eds.), *Beyond decoding: The behavioral and biological foundations of reading comprehension*. The Guilford Press.

Protopapas, A., Altani, A., & Georgiou, G. K. (2013). Development of serial processing in reading and rapid naming. *Journal of Experimental Child Psychology*, *116*(4), 914–929. https://doi.org/10.1016/j.jecp.2013.08.004

Protopapas, A., Katopodi, K., Altani, A., & Georgiou, G. K. (2018). Word Reading Fluency as a Serial Naming Task. *Scientific Studies of Reading*, *22*(3), 248–263. https://doi.org/10.1080/10888438.2018.1430804

Purpura, D. J., Hume, L. E., Sims, D. M., & Lonigan, C. J. (2011). Early literacy and early numeracy: The value of including early literacy skills in the prediction of numeracy development. *Journal of Experimental Child Psychology*, *110*(4), 647–658. https://doi.org/10.1016/j.jecp.2011.07.004

Rayner, K., & Slattery, Timothy. J. (2009). Eye Movements and Moment-to-Moment Comprehension Processes in Reading. In R. K. Wagner, C. Schatschneider, & C. Phythian-Sence (Eds.), *Beyond decoding: The behavioral and biological foundations of reading comprehension*. The Guilford Press.

Rinne, L. F., Ye, A., & Jordan, N. C. (2020). Development of arithmetic fluency: A direct effect of reading fluency? *Journal of Educational Psychology*, *112*(1), 110–130. https://doi.org/10.1037/edu0000362

Rosseel, Y. (2018). *lavaan—Latent Variable Analysis* [Computer software]. https://cran.r-project.org/package=lavaan

RStudio Team. (2021). *RStudio: Integrated Development Environment for R.* (1.4.1106) [Computer software]. RStudio, PBC. http://www.rstudio.com

Schwanenflugel, P. J., & Knapp, N. F. (2016). *The psychology of reading: Theory and applications*. The Guilford Press.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an 'outlier' orthography. *Psychological Bulletin*, *134*(4), 584–615. https://doi.org/10.1037/0033-2909.134.4.584

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. Oxford University Press.

Siemann, J. (2018). Evaluation of the Triple Code Model of numerical processing—Reviewing past neuroimaging and clinical findings. *Research in Developmental Disabilities*, 12.

Song, S., Georgiou, G. K., Su, M., & Hua, S. (2016). How Well Do Phonological Awareness and Rapid Automatized Naming Correlate With Chinese Reading Accuracy and Fluency? A Meta-Analysis. *Scientific Studies of Reading*, *20*(2), 99–123. https://doi.org/10.1080/10888438.2015.1088543

Statistics Norway. (2020a). *Educational attainment of the population*. Ssb.No. https://www.ssb.no/en/utdanning/statistikker/utniv/aar/2020-06-19

Statistics Norway. (2020b). *Number of Employments and Earnings*. Ssb.No. https://www.ssb.no/en/utdanning/statistikker/utniv/aar/2020-06-19

The jamovi project. (2021). *Jamovi* (version 1.6) [Computer software]. https://www.jamovi.org

Thomas, K. (1987). Numeracy in Early Modern England. *The Prothero Lecture*. *Transactions of the Royal Historical Society*, *37*, 103–132. https://doi.org/10.2307/3679153

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of Phonological Awareness and Rapid Automatic Naming Ability to the Growth of Word-Reading Skills in Second-to Fifth-Grade Children. *Scientific Studies of Reading*, *1*(2), 161–185. https://doi.org/10.1207/s1532799xssr0102_4

Torkildsen, J. von K., Arciuli, J., & Wie, O. B. (2019). Individual differences in statistical learning predict children's reading ability in a semi-transparent orthography. *Learning and Individual Differences*, *69*, 60–68. https://doi.org/10.1016/j.lindif.2018.11.003

Vaessen, A., & Blomert, L. (2010). Long-term cognitive dynamics of fluent reading development. *Journal of Experimental Child Psychology*, *105*(3), 213–231. https://doi.org/10.1016/j.jecp.2009.11.005

Vaessen, A., Gerretsen, P., & Blomert, L. (2009). Naming problems do not reflect a second independent core deficit in dyslexia: Double deficits explored. *Journal of Experimental Child Psychology*, *103*(2), 202–221. https://doi.org/10.1016/j.jecp.2008.12.004

Vanbinst, K., Ceulemans, E., Ghesquière, P., & De Smedt, B. (2015). Profiles of children's arithmetic fact development: A model-based clustering approach. *Journal of Experimental Child Psychology*, *133*, 29–46. https://doi.org/10.1016/j.jecp.2015.01.003

Vanbinst, K., Ghesquière, P., & De Smedt, B. (2012). Numerical Magnitude Representations and Individual Differences in Children's Arithmetic Strategy Use. *Mind, Brain, and Education*, *6*(3), 129–136. https://doi.org/10.1111/j.1751-228X.2012.01148.x

Vander Stappen, C., & Reybroeck, M. V. (2018). Phonological Awareness and Rapid Automatized Naming Are Independent Phonological Competencies With Specific Impacts on Word Reading and Spelling: An Intervention Study. *Frontiers in Psychology*, *9*, 320. https://doi.org/10.3389/fpsyg.2018.00320

Wang, Y., Ye, X., & Deng, C. (2020). Exploring mechanisms of rapid automatized naming to arithmetic skills in Chinese primary schoolers. *Psychology in the Schools*, *57*(4), 556–571. https://doi.org/10.1002/pits.22349

Willcutt, E. G., Petrill, S. A., Wu, S., Boada, R., DeFries, J. C., Olson, R. K., & Pennington, B. F. (2013). Comorbidity Between Reading Disability and Math Disability: Concurrent Psychopathology, Functional Impairment, and Neuropsychological Functioning. *Journal of Learning Disabilities*, *46*(6), 500–516. https://doi.org/10.1177/0022219413477476

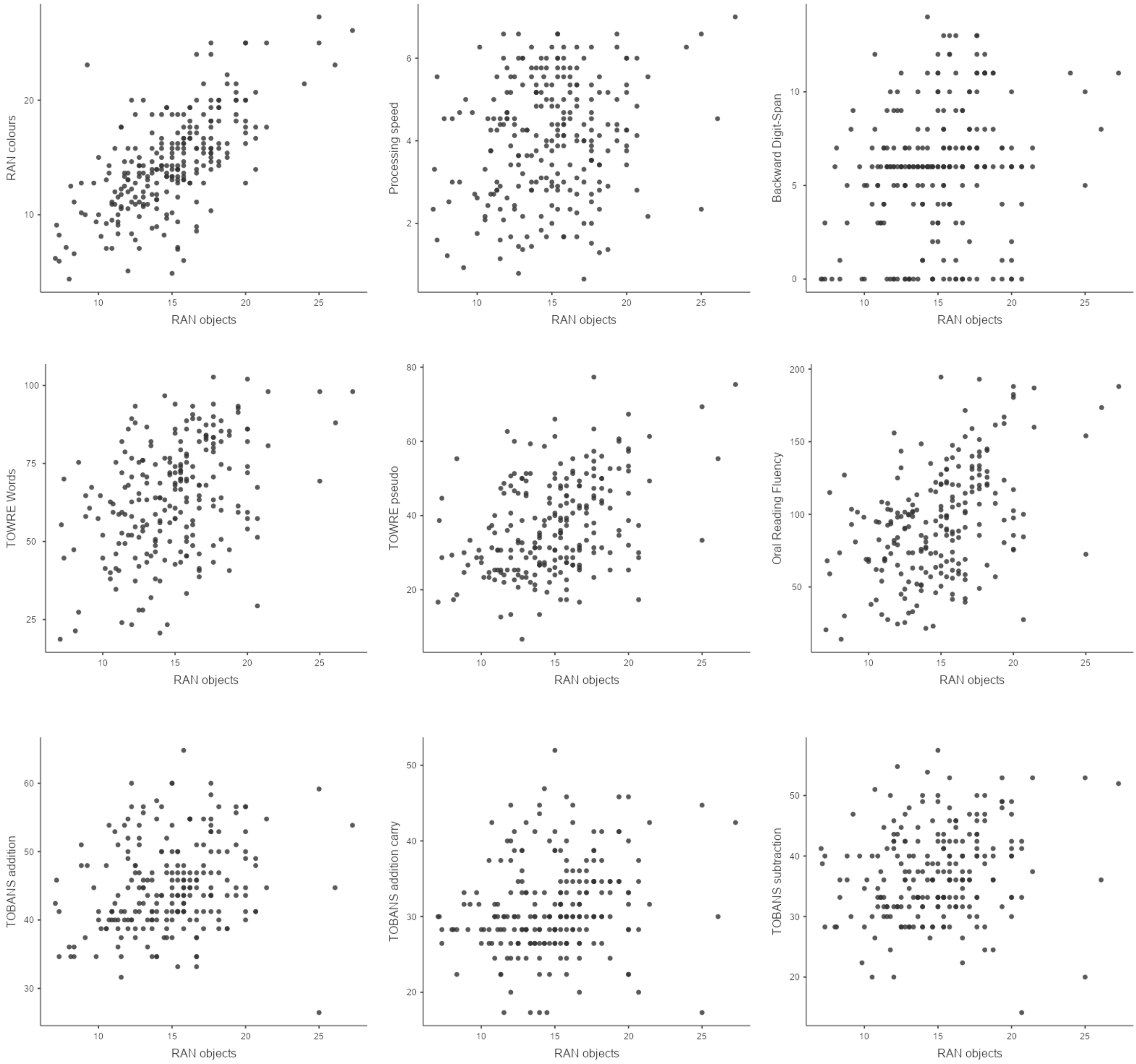
Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, *91*(3), 415–438. https://doi.org/10.1037/0022-0663.91.3.415

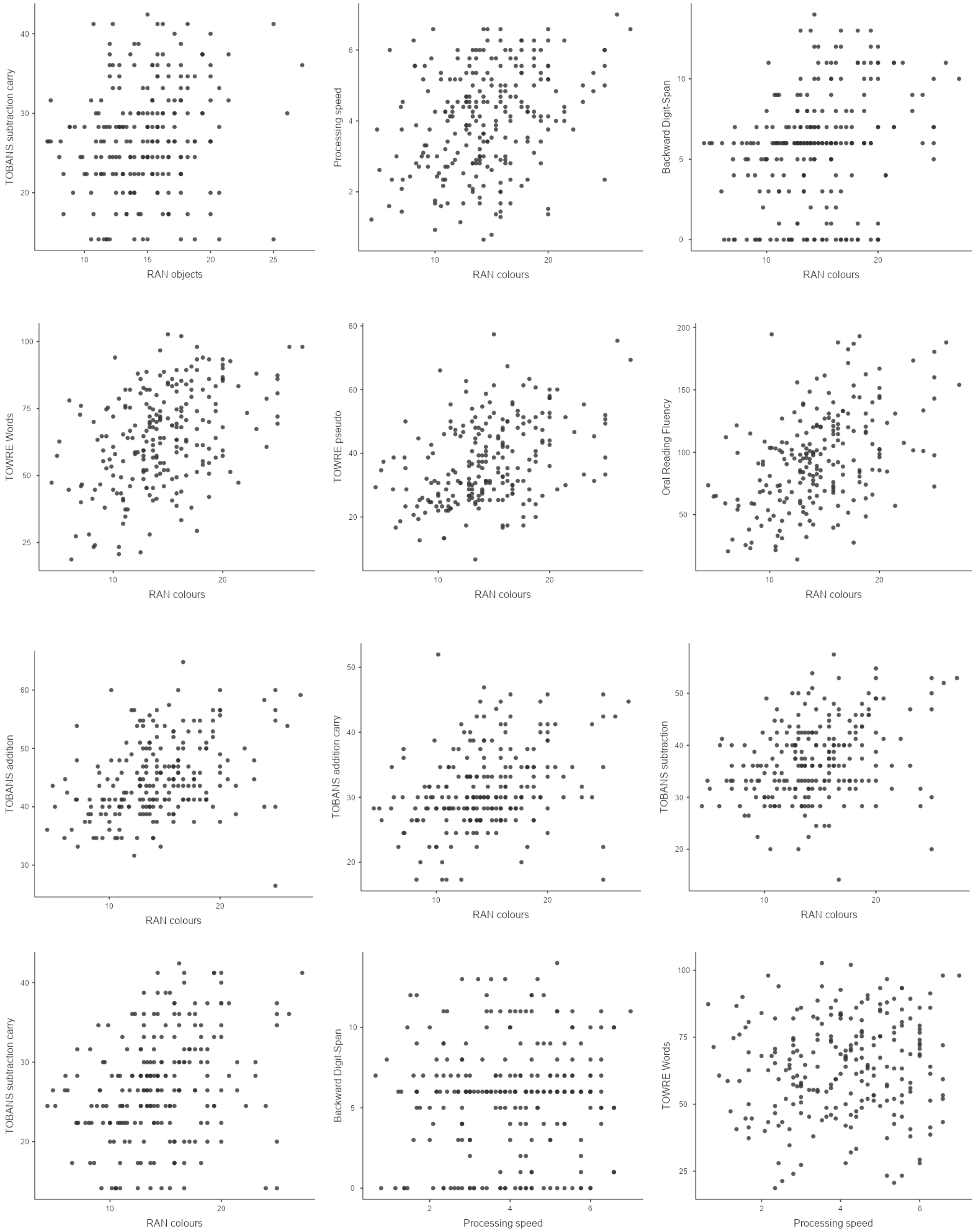
Ziegler, J. C., & Goswami, U. (2005). Reading Acquisition, Developmental Dyslexia, and Skilled Reading Across Languages: A Psycholinguistic Grain Size Theory. *Psychological Bulletin*, *131*(1), 3. https://doi.org/10.1037/0033-2909.131.1.3

14 Appendix

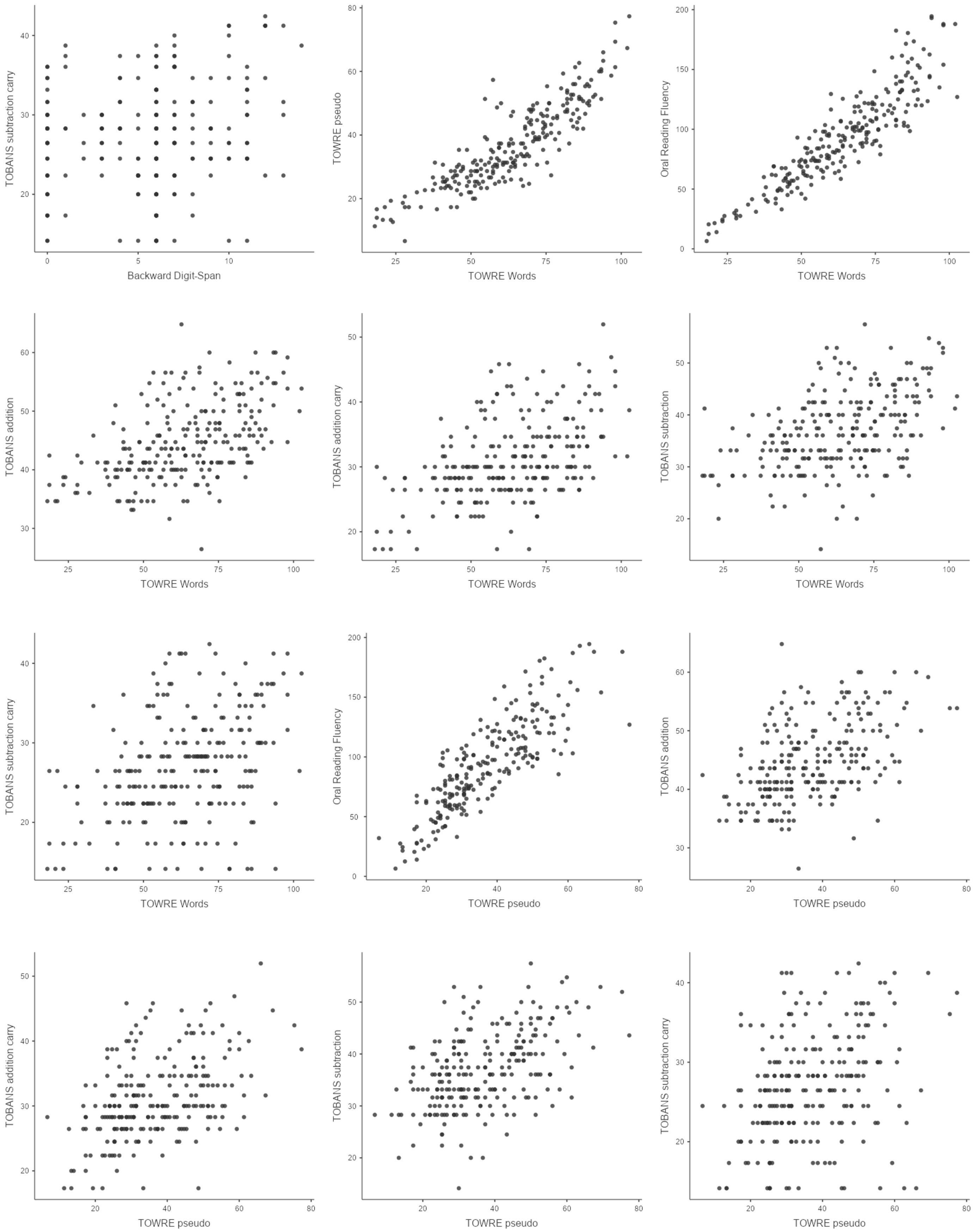
Figures 13-68: Bivariate Scatterplots

Note: Where relevant, variables are presented post-transformation




TOBANS subtraction carry
RAN objects
Processing speed
RAN colours
Backward Digit-Span
RAN colours
TOWRE Words
RAN colours
TOWRE pseudo
RAN colours
Oral Reading Fluency
RAN colours
TOBANS addition
RAN colours
TOBANS addition carry
RAN colours
TOBANS subtraction
RAN colours
TOBANS subtraction carry
RAN colours
Backward Digit-Span
Processing speed
TOWRE Words
Processing speed


TOWRE pseudo
Processing speed
Oral Reading Fluency
Processing speed
TOBANS addition
Processing speed
TOBANS addition carry
Processing speed
TOBANS subtraction
Processing speed
TOBANS subtraction carry
Processing speed
TOWRE Words
Backward Digit-Span
TOWRE pseudo
Backward Digit-Span
Oral Reading Fluency
Backward Digit-Span
TOBANS addition
Backward Digit-Span
TOBANS addition carry
Backward Digit-Span
TOBANS subtraction
Backward Digit-Span


TOBANS subtraction carry
Backward Digit-Span
TOWRE pseudo
TOWRE Words
Oral Reading Fluency
TOWRE Words
TOBANS addition
TOWRE Words
TOBANS addition carry
TOWRE Words
TOBANS subtraction
TOWRE Words
TOBANS subtraction carry
TOWRE Words
Oral Reading Fluency
TOWRE pseudo
TOBANS addition
TOWRE pseudo
TOBANS addition carry
TOWRE pseudo
TOBANS subtraction
TOWRE pseudo
TOBANS subtraction carry
TOWRE pseudo

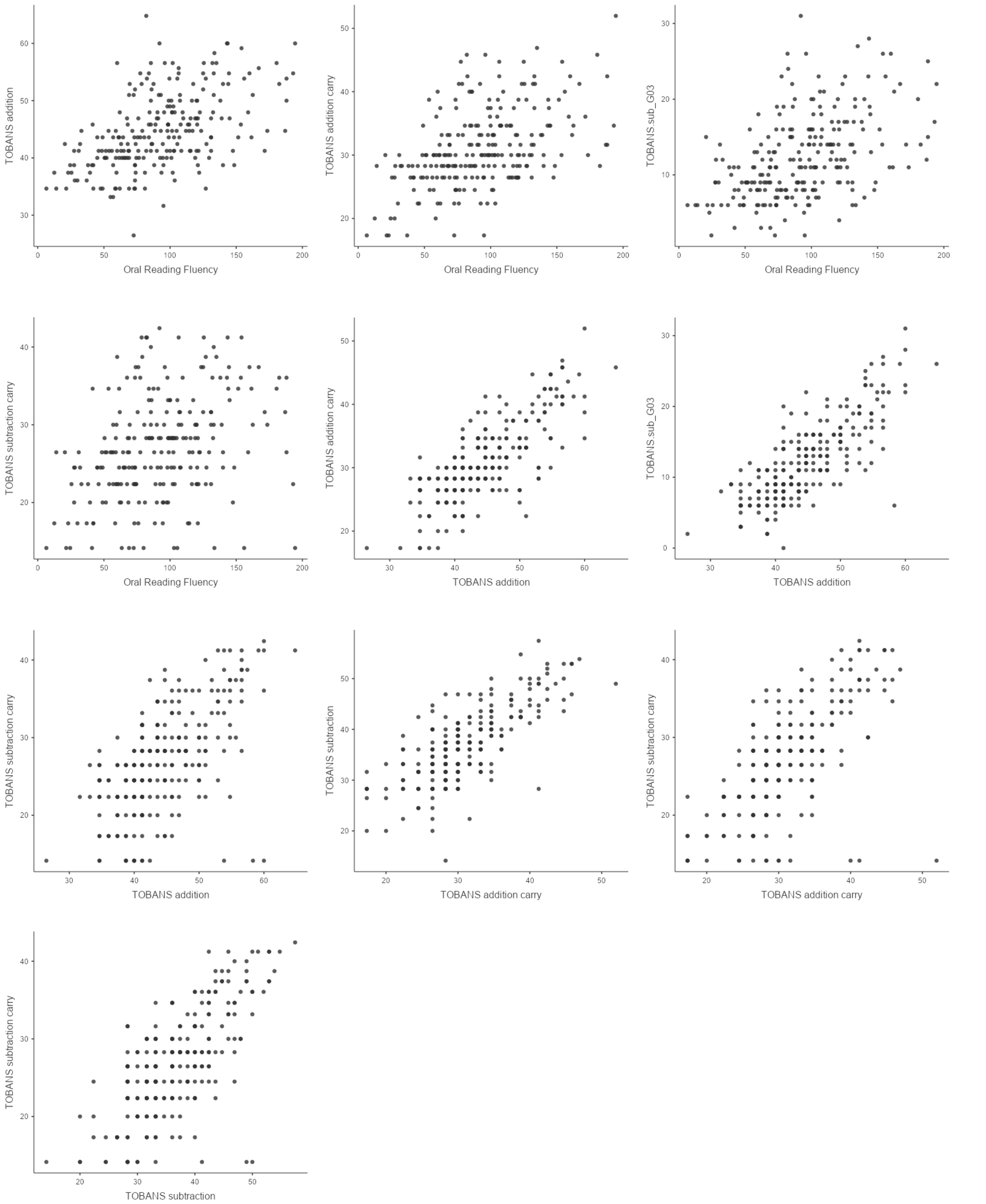

TOBANS addition
Oral Reading Fluency
TOBANS addition carry
Oral Reading Fluency
TOBANS.sub_G03
Oral Reading Fluency
TOBANS subtraction carry
Oral Reading Fluency
TOBANS addition carry
TOBANS addition
TOBANS.sub_G03
TOBANS addition
TOBANS subtraction carry
TOBANS addition
TOBANS subtraction
TOBANS addition carry
TOBANS subtraction carry
TOBANS addition carry
TOBANS subtraction carry
TOBANS subtraction

Table 4: Parameter estimates of the covariance model

Variables	*Estimate*	*Standard Error*	*Standardized Estimate*
Regressions			
Cov ~ RAN	2.873	0.383	0.638
Reading Fluency ~ Processing Speed	-2.139	0.715	-0.173
Arithmetic Fluency ~ Working Memory	0.271	0.103	0.162
Correlations			
Subtraction ~~ Subtraction with carry	3.148	1.565	0.189
RAN ~~ Processing Speed	1.754	0.406	0.328
RAN ~~ Working memory	4.052	0.966	0.314

Note. X ~ Y means X is regressed on Y. X ~~ Y means the correlation of X and Y

Table 5: Correlation residuals of the covariance model

	RAN objects	RAN colours	Backward digit-span	Processing speed	TOWRE words	TOWRE pseudo	Oral reading fluency	TOBANS addition	TOBANS addition with carry	TOBANS subtraction	TOBANS subtraction with carry
RAN objects	0.006										
RAN colours	0.007	0.007									
Backward digit-span	-0.014	0.016	0.000								
Processing speed	-0.035	0.025	0.048	0.000							
TOWRE words	0.018	-0.032	0.035	0.046	-0.011						
TOWRE pseudo	0.033	0.029	0.024	0.020	-0.007	-0.010					
Oral reading fluency	0.056	0.017	0.077	0.050	-0.010	-0.019	-0.010				
TOBANS addition	-0.029	0.043	0.014	0.101	0.001	0.031	0.037	0.008			
TOBANS addition with carry	-0.062	-0.008	0.040	0.038	-0.027	-0.010	-0.002	0.006	0.008		
TOBANS subtraction	-0.051	0.003	0.027	0.116	-0.031	0.033	0.000	0.012	0.003	0.008	
TOBANS subtraction with carry	-0.010	0.040	-0.004	0.135	-0.034	-0.036	-0.037	-0.026	0.049	0.006	0.005