Pre-school predictors of early arithmetic skills: a two year longitudinal study

Stefanie Habermann

Thesis submitted for a PhD qualification

Supervisors:
Prof. Chris Donlan
Prof. Charles Hulme
Dr Alexandra Perovic

UCL, UK
Department of Psychology and Language Sciences

January 2018

Author's Declaration

I, Stefanie Habermann confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. This research was supported by an Economic and Social Research Council (ESRC) studentship awarded to the candidate.

Selected aspects of the research have been presented elsewhere:

Habermann, S., Donlan, C., Göbel, S. and Hulme, C. Pre-school predictors of early arithmetic skills: a two year longitudinal study. In preparation.

Habermann, S., Donlan, C., Göbel, S. and Hulme, C. (2017, February). *Pre-school Predictors of 6 Year Olds' Arithmetic Skills: The Importance of Knowing Your Numbers.* Poster presented at the Royal Society scientific meeting 'The origins of numerical abilities', London, UK.

Habermann, S., Donlan, C., Göbel, S. and Hulme, C. *The Importance of Pre-school Transcoding Skills in the Development of Arithmetic.* Oral presentation at Experimental Psychology Society Annual Meeting, London, UK, 5th January, 2017.

Habermann, S., Donlan, C., Göbel, S. and Hulme, C. (2016, September). *Pre-school Predictors of 6 Year Olds' Arithmetic Skills: The Importance of Knowing Your Numbers.* Poster presented at the 'Domain-General and Domain-Specific Foundation of Numerical and Arithmetic Processing' workshop, Tübingen, Germany.

Some of the research presented here was conducted in collaboration with others:

The research in Chapter 4 and 5 was aided by Kate Murray who collected the data at Time 4 as part of her BSc Psychology and Language Sciences dissertation project.

The magnitude comparison tasks used in this thesis were initially developed by Sarah Watson in her PhD project led by Silke Göbel.

Abstract

The purpose of this thesis was to examine longitudinal predictors of children's early arithmetic abilities with a particular focus on the relation to the approximate number system (ANS), language and numeracy skills as well as background measures of cognitive abilities. This longitudinal study assessed children five times over a 25-month period beginning in nursery classes and continuing to the end of Year One (the first complete year of formal schooling). The thesis investigated the concurrent and longitudinal predictive importance of ANS, numeracy, language and cognitive abilities in children's arithmetic development using structural equation modelling. Path models found different concurrent predictors of arithmetic at each time point and only transcoding, the ability to translate between the verbal number code and the Arabic numeral, was a consistently recurring predictor. Furthermore, children's nonverbal intelligence and their understanding of language specific to mathematics related significantly to early arithmetic (pre-school) whereas children's magnitude comparison skills were significantly associated with arithmetic scores in Year One. The longitudinal analysis showed that transcoding was the only unique predictor of arithmetic and neither ANS nor language and cognitive skills were significant independent contributors to the prediction of children's arithmetic abilities 25 months later.

Table of contents

List of Tables

List of Figures

List of Appendices

Acknowledgements

I would like to specially thank my supervisors, Chris Donlan for his support, guidance and encouragement throughout my PhD and Charles Hulme for his time and patience teaching me to use MPlus and his guidance with the data analysis. I would also like to acknowledge Alexandra Perovic for the insightful discussions on all matters linguistics.

I would like to extend a very special thank you to the Tudor Court Primary school in Chafford Hundred for inviting me to work with them. A big thanks to all the children (and parents) who participated in this research and the teachers for their advice and warm welcome. I am grateful to the ESRC for funding this research. I would also like to say thank you to all the undergraduate students for helping me carrying out the data collection.

Most importantly, I would like to say thank you to my family and friends for their continued support. I am indebted to my husband Michael who believed in me and helped me keep things in perspective and I deeply appreciate all your love and support throughout this PhD. I owe a big thank you to a very special person, my daughter Lucy, for the patience she showed during my thesis writing. You are the best thing that happened to me!

Chapter 1. Background

Competence in numeracy is of great and increasing importance for success in modern society. The development and understanding of mathematics is a complex combination of factual knowledge (memorised number facts), procedural knowledge (understanding of how to proceed in order to get an answer) and conceptual knowledge (understanding why a strategy works and is more effective than another) (Dowker, 2005). Proficiency in arithmetic is one of the most crucial achievements in primary education. Thereby children have to master the basic principles of counting and arithmetic concepts (Desoete and Grégoire, 2006; Nunes and Bryant, 1996; Gelman and Gallistel, 1978).

Despite a growing literature mapping the development of mathematical skills, little is known about longitudinal predictors of early, arithmetic skills. The search for cognitive-developmental precursors of basic arithmetic has focussed on working memory (e.g. Bull, Epsy, and Wiebe, 2008), counting (e.g. Donlan, Cowan, Newton, and Lloyd, 2007), language (Kleemans, Segers, and Verhoeven, 2011; 2012), number knowledge (Jordan, Kaplan, Ramineni and Locuniak, 2009) and magnitude processing (Mazzocco, Feigenson and Halberda, 2011; Durand, Hulme, Larking, and Snowling, 2005). The interaction of these predictors has received little attention. Language is the core medium of instruction in school and therefore crucial for the acquisition of knowledge and skills across the curriculum. On the other hand, there is strong intuitive appeal in a modular developmental model in which infants' ability to identify numerical differences between nonsymbolic stimulus sets (e.g. arrays of dots) provides the basis on which symbolic number processing is constructed, in turn providing the semantic framework needed for the development of arithmetic skills. A central aim of this thesis is to establish whether magnitude comparison tasks are reliable longitudinal predictors of early arithmetic performance, once number knowledge, language and cognitive skills are taken into account.

1.1 Models of numerical processing.

1.1.1 General models of numerical processing.

Several models have been proposed to explain adults' ability to solve arithmetic problems. The most notable models are the Triple Code Model (Dehaene, 1992), McCloskey's Modular Model (McCloskey, Sokol and Goodman, 1986), MATHNET (McCloskey and Lindemann, 1992) and the Network Interference Model (Campbell, 1995).

1.1.1.1 McCloskey's Modular Model and MATHNET.

McCloskey and colleagues proposed the modular model (Figure 1.1) after studying adults with acquired dyscalculia (McCloskey and Lindemann, 1992; McCloskey et al., 1985; 1986). The model incorporates distinct subsystems for particular arithmetic abilities. In particular, the model distinguishes between calculation and number processing which in itself is divided into number comprehension and production. Verbal and Arabic numerals are dealt with separately. The calculation system involves three mechanisms: one for operation processing, one for procedures and one for fact storage and access. The first is used when processing symbols and words related to symbols (e.g. plus) whereas the second mechanism involves rules for the processing of operations such as addition, subtraction, multiplication and division. Storage and retrieval of facts rely on the facts mechanism. The semantic processing of numbers is a key component of the Modular Model. It converts input into abstract semantic representations, similar to the magnitude representations posited by Dehaene (1992), which then are processed in the calculation system and/or output through the numeral production system. For example, the semantic form of the number '123' is {1} 10^2, {2} 10^1 and {3} 10^0 whereas {1}, {2} and {3} represent quantities and 10^n is the power of ten.

According to the model, a problem needs to be converted into a semantic representation before it can activate stored facts in memory. The semantic representation of the answer can then be accessed and converted into the appropriate format for the response. Therefore, an impairment in the semantic representation system impedes the performance on all arithmetic problems.

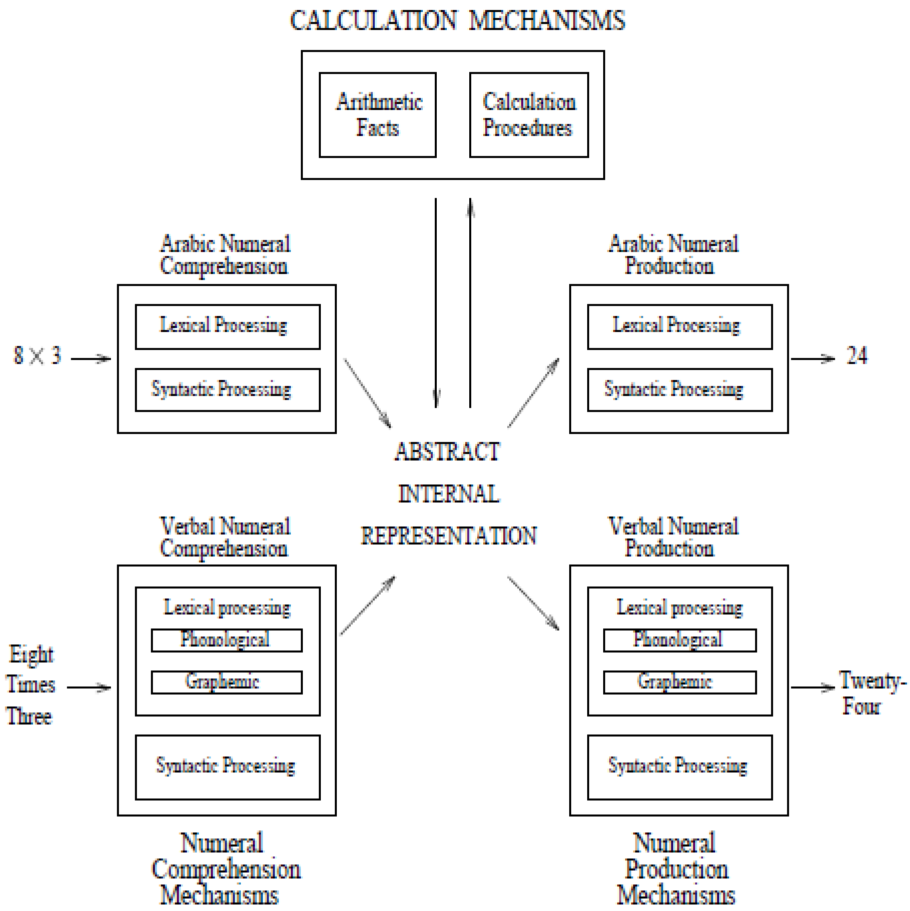


Figure 1.1. Modular Model proposed by McCloskey, Caramazza, and Basili (1985). From Edelman, Abdi, and Valentin (1996, p.48).

McCloskey and Lindemann, (1992) proposed the MATHNET model (Figure 1.2) with three layers. The first layer consists of 26 input units which are connected to the 40 hidden units, layer two, which are connected to the 24 answer units. A particular feature of this model is that the answer units are also interconnected. All connections are symmetric and bidirectional and the hidden and answer units tend towards either a positive or negative activation. However, the problem arises that the only solution method for an arithmetic problem is retrieval from memory neglecting to explain non-retrieval solutions such as transformation or counting-based strategies (Geary and Wiley, 1991; LeFevre, Sadesky and Bisanz, 1996).

1.1.1.2 Dehaene's Triple Code Model.

The most influential model of adult numerical processing is Dehaene's Triple Code Model (Dehaene, 1992, Figure 1.3). The model proposes that numbers are represented mentally in three different codes: the auditory verbal code or auditory

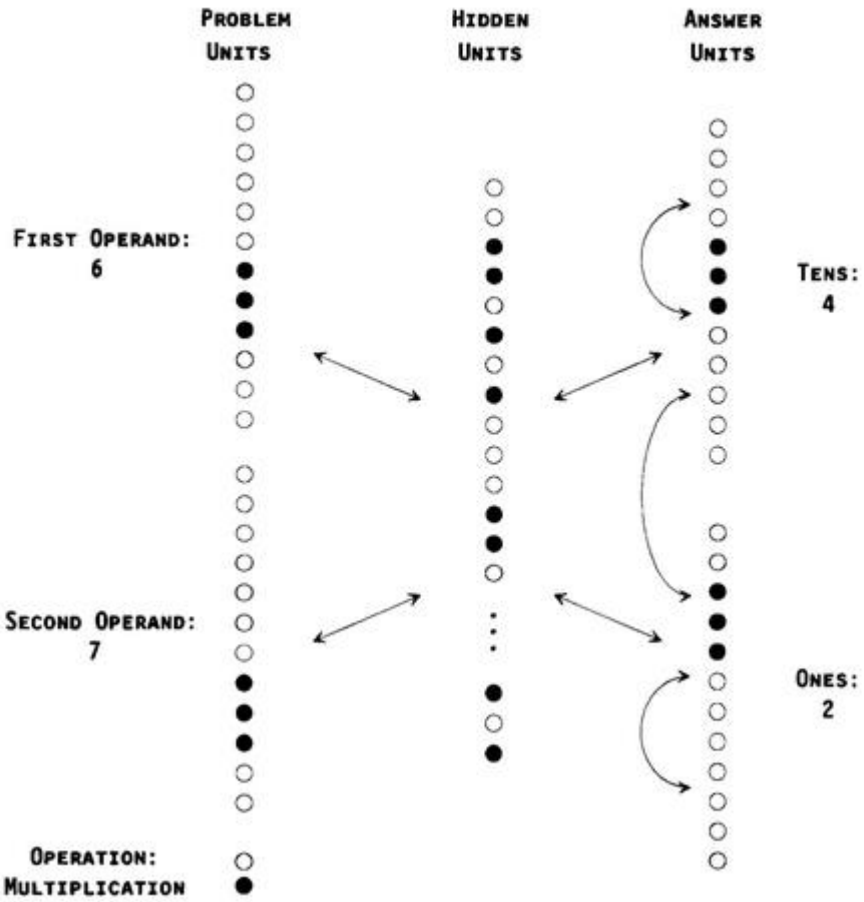


Figure 1.2. MATHNET model proposed by McCloskey, and Lindemann (1992).

verbal word frame, the visual Arabic number form, and the analogue magnitude code. The codes are interlinked via pathways that translate from one code to another and the task being performed determines which form of mental number representation is used. The auditory verbal code draws on general language processing systems and deals with tasks such as verbal counting and multiplication tables. The visual Arabic number code is created and manipulated using Arabic numerals and is used in multi-digit calculations. These two codes are considered to be unique to humans.

The analogue magnitude code in which quantity or magnitude is represented in an approximate way is believed to be shared by animals and humans, even preverbal infants alike (Dehaene, 1992; Whalen, Gallistel and Gelman, 1999; Wynn, 1998; Rugani, Vallortigara, Priftis and Regolin, 2015) and often referred to as the number sense (Dehaene, 1997). The analogue magnitude code is hypothesised to take the form of a compressed number line following Weber's Law. Magnitude representations are supposed to be compressively spaced with larger numbers being less discriminable than smaller numbers.

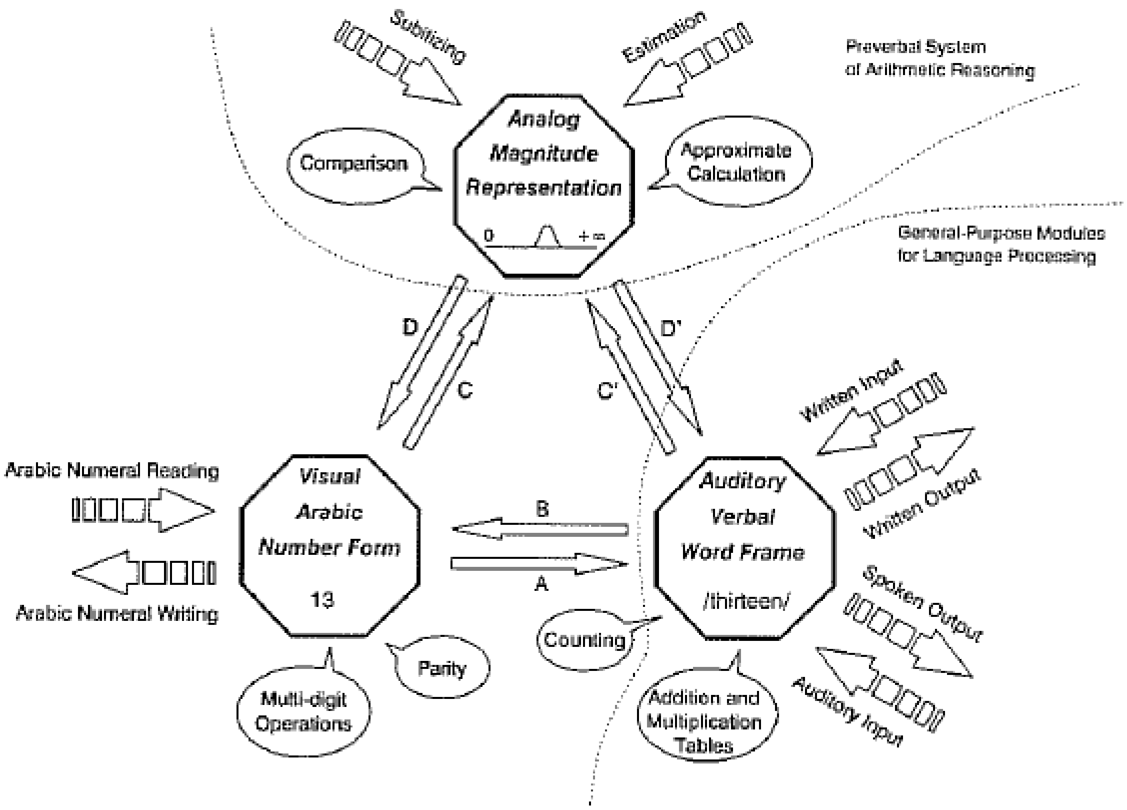


Figure 1.3. Dehaene's Triple Code Model (Dehaene, 1992, p. 31). Three representations are depicted as octagons. Large arrows indicate input-output processes, whereas thin arrows depict internal translation processes.

Dehaene also proposed two separate pathways for solving arithmetic problems, the direct asemantic route for over learned calculations and the indirect semantic route for exact arithmetic. The former relies on activation of the auditory verbal code because facts are stored as a "learned lexicon of verbal associations" (Dehaene, 1992, p.34) in memory. The latter involves the analogue magnitude code and is proposed to be used in subtractions and more complex addition problems.

1.1.1.3 Network interference model.

The Network interference model (Campbell, 1995, Figure 1.4) proposes multiple internal codes. Similar to Dehaene (1992), codes include the magnitude, verbal and visual code for numbers. Encoding and calculation processes are assumed to be interactive which explains why large mathematical problems in word format take longer to solve. It further hypothesises that larger magnitudes are less discriminable than smaller quantities as proposed in the Triple Code model (Dehaene, 1992). This encoding complex model posits that numbers evoke “an integrated network of format-specific number codes and processes that collectively mediate number comprehension, calculation, and production, without the assumption of central representation” (Campbell and Clark, 1988, p. 204). According to the model, additions and comparisons depend on qualitatively different processes (Takahashi and Green, 1983). The model, however, does not account for non-retrieval solution methods.

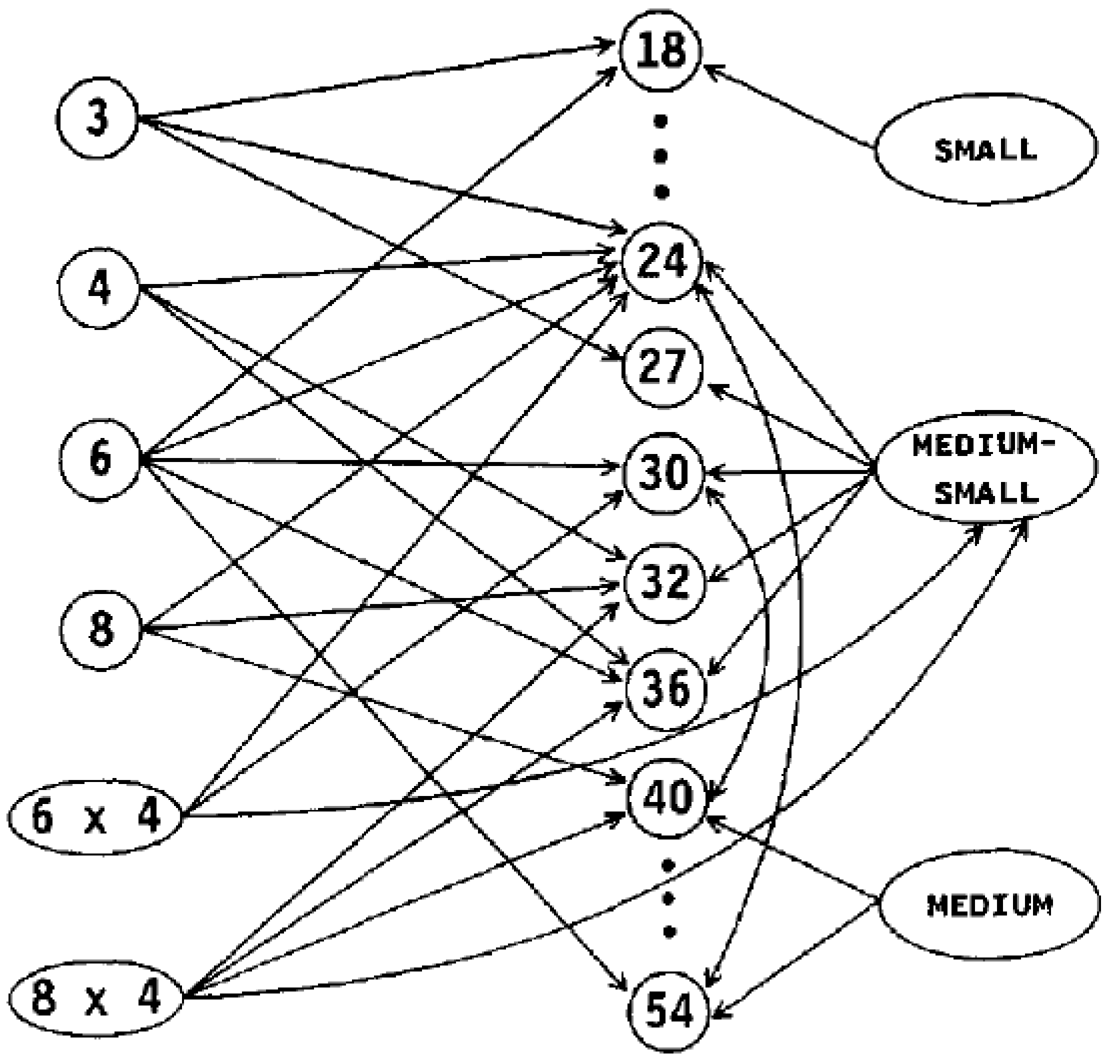


Figure 1.4. Schematic shows some of the nodes and connections described in the Network Interference Model (Campbell, 1995).

1.1.2 Developmental models of arithmetic.

Most developmental models of arithmetic are models of children's problem solution and were inspired by the adult models. Models discussed in this thesis are the Adaptive Strategy Choice Model (ASCM, Siegler and Shipley, 1995), the Four-Step Developmental Model of Numerical Cognition (von Aster and Shaley, 2007), the Pathways to Mathematics Model (LeFevre, Fast, Skwarchuk, Smith-Chant, Bisanz, Kamawar and Penner-Wilger, 2010) and the The Integrated Theory of Numerical Development (Siegler and Braithwaite, 2017).

1.1.2.1 Adaptive Strategy Choice Model.

Siegler and Shipley's (1995) Adaptive Strategy Choice model (ASCM, Figure 1.5) is a "computer simulation of how strategy choices are made and how they change with age and experience" (p.72). It aims to explain how and why people choose a specific strategy among alternative strategies to perform fast and accurately. ASCM's three assumptions are: First, a database maintains information on performance and outcome of a particular strategy which plays an important role in choosing strategies in the future. This database is dynamic and updates the information after the strategy was used and it also stores information on global data (average speed and accuracy for a particular strategy), feature data (speed and accuracy for each strategy on problems with a particular structural feature) and local data (speed and accuracy of the strategy on particular problems). Second, choosing a strategy is determined by past performance of the strategy and the predicted performance of the strategy if that strategy would be used. Third, strategy choice depends on the each strategy's strength (speed and accuracy) in comparison with alternative strategies. New strategies are developed by boosting the speed and accuracy information of novel strategies.

The model posits two distinct strategy-choice pathways for procedures and retrieval. If the former is activated, the procedure will be performed and the answer is produced. Contrarily, a successful retrieval depends on the strength of each strategy. Procedure is often chosen for problems were the association between a problem and an answer is weak, whereas the retrieval pathway is more likely to be chosen for problems with strong answer associations. A retrieval answer will only produce an answer if the confidence in the correctness of the answer exceeds a set criterion otherwise the models attempts another retrieval. The procedure pathway will then only be chosen after numerous failed retrieval attempts.

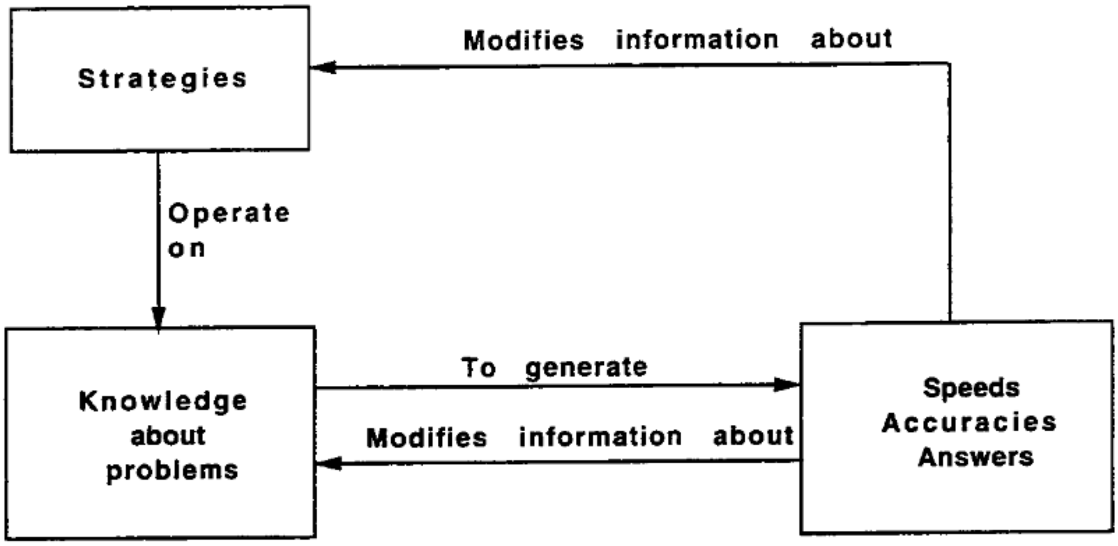


Figure 1.5. Overview of Adaptive Strategy Choice Model proposed by Siegler and Shipley (1995). From Siegler and Lemaire (1997), p. 73.

1.1.2.2 Four-Step Developmental Model of Numerical Cognition

The Four-Step Developmental Model of Numerical Cognition (von Aster and Shalev, 2007) describes four stages of numerical cognition in which children move through the stages as they progress in arithmetic competency through exposure and formal schooling and an increase in working memory capacity (Figure 1.6). The first stage consists of an inherited core-system of magnitude representation, similar to Dehaene’s number sense, which entails subitizing and approximation abilities. This basic meaning of number is a prerequisite for the acquisition of more complex mathematical skills. Pre-school children move on to the linguistic stage of numeracy (step 2) where children acquire the verbal number codes. In step 3, children learn the Arabic number system and the symbolic representations of magnitudes in school. Typical mathematical skills developing at this stage are written calculations and odd-even decisions. The final stage, the mental number line, develops during school years as children acquire the concept of ordinality, a second core principal of number.

Von Aster and Shalev (2007) further propose that failure to establish a stage appropriately may lead to developmental delays in acquiring the follow-on stages or dyscalculia. For example, a child that has an inappropriate concept of magnitude (step 1) because of genetic vulnerability may still learn the verbal codes (stage 2) by rote memory, but the codes are void of the meaning putting the child at risk of pure developmental dyscalculia.

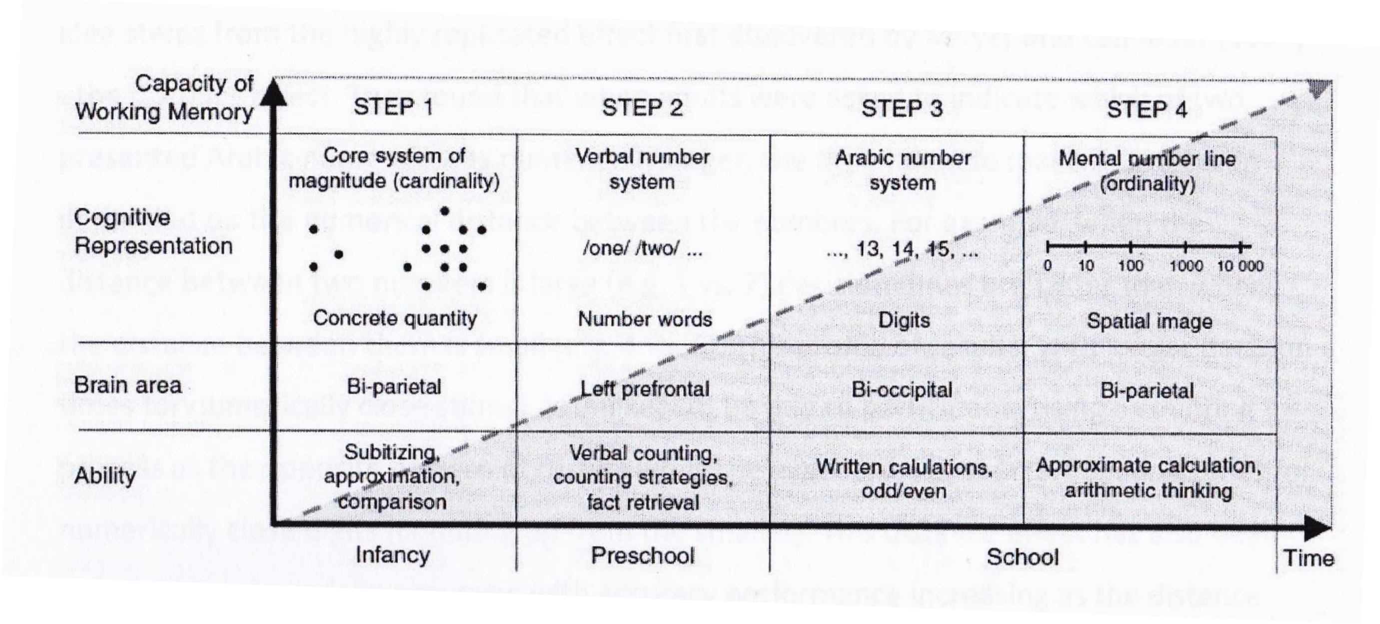


Figure 1.6. Four-step-developmental model of numerical cognition. From Von Aster and Shalev (2007), p. 870. Shaded area below line represents increasing working memory.

1.1.2.3 Pathways to Mathematics

LeFevre and colleagues postulate the Pathways to Mathematics model (Figure 1.7) focusing on the relationships between children's mathematical skills and cognitive precursors, early numeracy skills and mathematical outcomes. (LeFevre et al., 2010). This model posits three separate pathways: quantitative, linguistic and spatial attentional. Each of these pathways contributes individually to the acquisition of early numeracy abilities. Furthermore, the models proposes that the linguistic, quantitative and spatial attention pathways vary in their contribution to mathematical performance depending on the demands of the arithmetic problem. According to the model, linguistic skills are linked to children’s symbolic number system knowledge. The second skill pathway comprises quantitative abilities and processing numerical magnitudes. Spatial attention forms a third pathway with connections across a variety of numerical and mathematical skills.

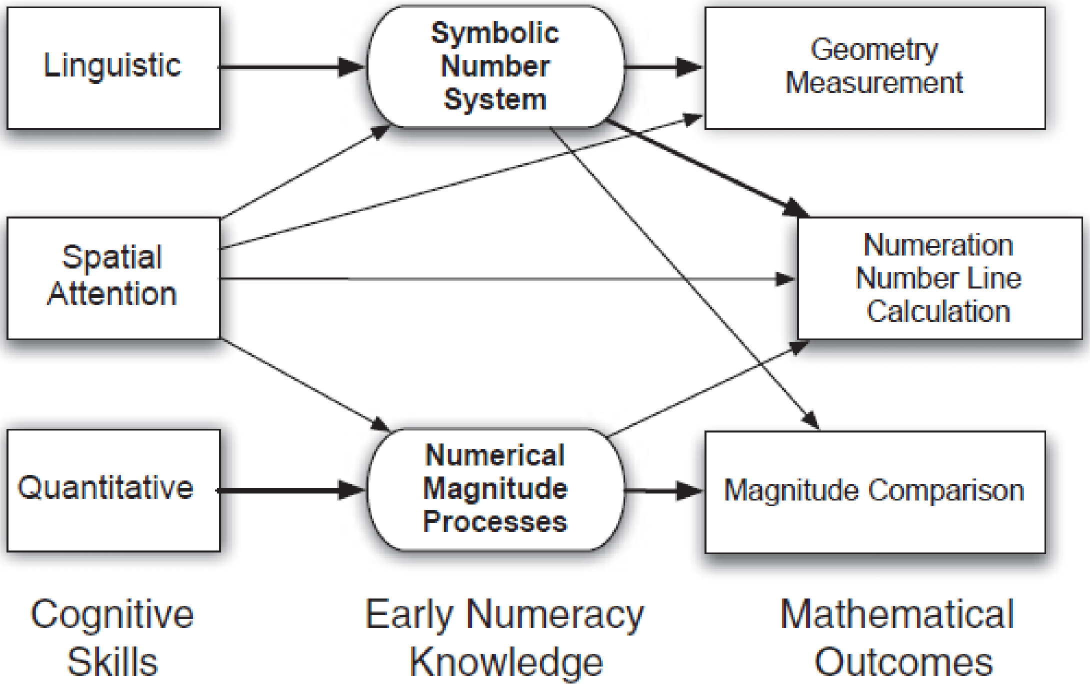


Figure 1.7. The Pathways to Mathematics model proposed by LeFevre et al. (2010), p. 1755.

Sowinski, LeFevre, Skwarchuk, Kamawar, Bisanz and Smith-Chant (2014) further expanded the quantitative pathway to include not only magnitude comparison but also counting and subitizing (ability to quickly and exactly enumerate small quantities; Clements, 1999). They also introduced a working memory pathway (the original model only focused on visuo-spatial attention). These new pathways were examined in relation to backward counting, arithmetic fluency, calculation, number system knowledge and reading. As expected, all three pathways contributed to backward counting and arithmetic fluency, but only the linguistic and quantitative pathway were uniquely predicting calculation and number system knowledge. Word reading was solely predicted by the linguistic pathway.

1.1.3 The Integrated Theory of Numerical Development

Siegler and Braithwaite (2017) proposed the integrated theory of numerical development assuming that the core of numerical development is an increase in understanding of numerical magnitudes. The theory posits five assumptions:

1. Magnitudes of numbers are represented on a mental number line in humans and animals. This number line is a dynamic structure that represents small numbers first and “then is progressively extended

rightward to include larger whole numbers, leftward to include negative numbers, and interstitially to include fractions and decimals" (Siegler and Braithwaite, 2017, p. 3). Figure 1.8 shows the approximate age range of the changes in the mental number line.

2. The representation of whole numbers shifts from a compressive, approximately logarithmic distribution towards a linear distribution. This shift occurs first for small whole numbers than larger whole numbers based on children's experience with the number range.
3. All real numbers can be presented as magnitudes on a mental number line and are thus ordered on the number line.
4. Whole and rational number knowledge is related and predictive of arithmetic attainment and more advanced aspects of mathematics.
5. Interventions to enhance numerical magnitudes knowledge positively affect arithmetic attainment.

The integrated theory presumes that numerical magnitudes are represented on a horizontally oriented mental number line, at least for many Western and Eastern cultures. Smaller numbers are presented on the left and larger number on the right of the number line. Rugani et al. (2015) found that newborn chicks spontaneously associate small numbers (e.g. "2") with the left and larger numbers (e.g. "8") with the right side when trained on the number "4". There is evidence for the mental number line representation of numbers provided by distance effects and the SNARC effect (spatial-numerical association of response codes). The former comes from the finding that the identification of the larger of two numbers is faster for numerically farther apart pairs of numbers (Moyer and Landauer, 1967). The SNARC effect explains the finding that responses for smaller numbers are faster when pressing a button on the left hand side and responses for larger numbers are faster when pressing a button on the right hand side (Dehaene, Dupoux and Mehler 1990).

Siegler and Braithwaite (2017) posit that numerical magnitude knowledge is related to and predictive of arithmetic development. Numerical magnitude knowledge is conventionally measured using the number-to-position task. Children are asked to indicate the position of a target number on a blank number line. Research has shown that older children perform better on number lines (as indicated by the difference between actual position and children's estimated position) than younger children, that performance on the number line is significantly associated with arithmetic skills (Siegler and Booth, 2004; Booth and Siegler, 2006, 2008) and that there is a shift from a logarithmic to a linear distribution of numerical magnitudes between five and eight years (Booth and Siegler, 2008; Siegler, Thompson and Opfer, 2009). However, these findings come mostly from cross-sectional studies.

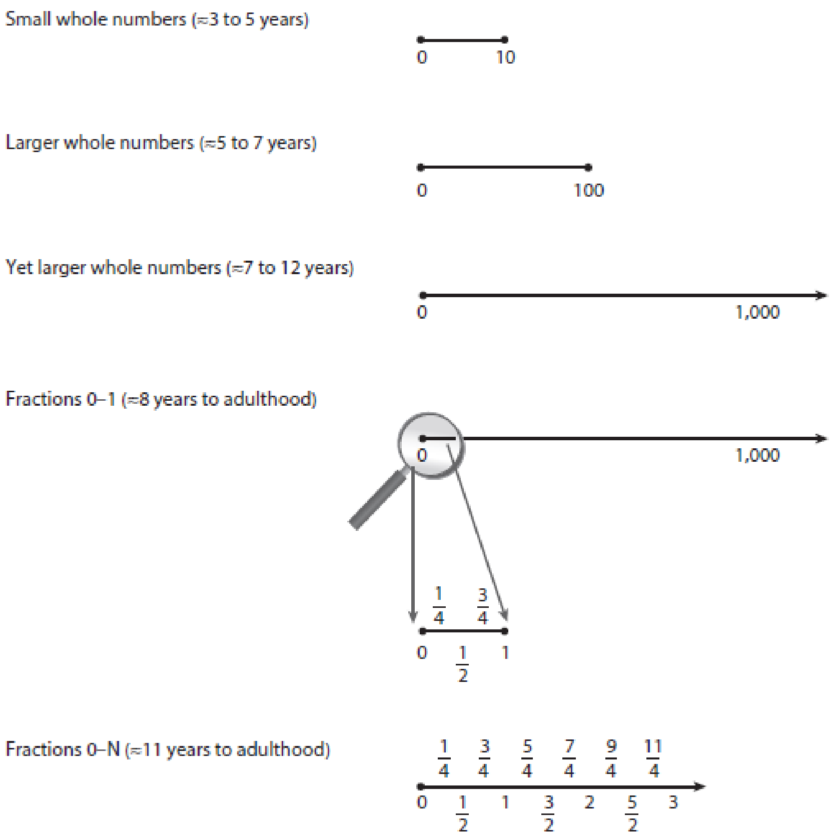


Figure 1.8. Mental number line model proposed by Siegler and Braithwaite (2017), p. 190, shows the approximate age ranges of major changes to the size and types of symbolic numbers whose magnitudes individuals can represent.

Some studies identify the involvement of language and counting alongside magnitude estimation as developmental associates of early arithmetic (Praet, Titeca, Ceulemans and Desoete, 2013). Indeed, Praet and colleagues (2013) explored the relationship between arithmetic and children’s estimation using number words, dots and Arabic numerals, adding language as a covariate (from kindergarten through to grade two). The results revealed that Arabic numerals were more linearly distributed than number words and that language explained kindergartener’s arithmetic performance, but not the growth of arithmetic. Children's untimed math performance was predicted by number line estimation.

Muldoon, Towse, Simms, Perra and Menzies (2013) assessed 5-year-olds over a 12 month period, with repeated measurement of number line estimation skills, counting ability and math achievement. They showed that counting was the largest contributor to children's math performance and only linear fit of number estimation on the 0-20 scale at 5 years and linear fit of number estimation on the 0-100 scale at six years made a significant contribution. Some studies propose that, rather than being a precursor of mathematical achievement, number line acuity and math performance both influence each other during development from pre-school through early school years (Friso-van den Bos, Kroesbergen, Van Luit, Xenidou-Dervou, Jonkman, Van der Schoot and Van Lieshout, 2014; LeFevre, Lira, Sowinski, Cankaya, Kamawar and Skwarchuk, 2013).

Evidence form cross-cultural research by the same team revealed that children's number estimations were related to some but not all mathematical skills in English and Chinese children. Although Chinese children are typically precocious when it comes to mathematical development, their estimations on the mental number line were not more linearly distributed or accurate than an older Western sample with equivalent math scores suggesting that linearity may not be a driver for math attainment. It also emerged that young children display numbers as accurately in the vertical as the horizontal orientation (Simms, Muldoon and Towse, 2013).

Ramini, Siegler and Hitti (2012) found that playing a linear number board game improved low-SES children's number line estimation, magnitude comparison, numeral identification and counting skills. Further evidence comes from an intervention study examining extremely preterm-born (EP) children (Simms, Gilmore, Cragg, Marlow, Wolke and Johnson, 2012). The authors compared EP children's performance on cognitive tests and number line estimation tasks to term-born control children. They reported that EP children performed worse than the controls in all tests, but different relationships between mathematical attainment and number estimation were found in the two groups. The relationship between number estimation and mathematical achievement was stronger in EP children and remained significant after controlling for cognitive abilities only in the EP children. The authors conclude that math attainment in EP children was associated with accuracy of numerical representations and not their general cognitive abilities.

1.2. Mathematics and Language.

Despite a growing literature mapping the development of mathematical skills, the interaction of language and arithmetic concept formation has received little attention. However, language is the core medium of instruction in school and therefore crucial for the acquisition of knowledge and skills across the curriculum. For this reason alone, it is essential to look at the role of language in children's early arithmetic skills.

Motivation for exploring the relationship between language and numeracy derives from findings suggesting that the development of numeracy is critically dependent on linguistic representations. Early support for this view came from scholarly examination of the cross-linguistic underpinnings of number systems (Hurford, 1987). More recently, an important perspective has been offered by studies assessing the mathematical skills of children with specific language impairment (SLI) in order to determine the impact of an impaired linguistic system on numeracy development. Children with SLI have impaired linguistic abilities which are not caused by hearing loss, physical disabilities or environmental influences. However, their nonverbal intelligence seems to be within normal range (APA, 1994).

Studies indicate that SLI may affect a wide range of numeracy skills differently (Donlan, Bishop and Hitch, 1998; Donlan, and Gourlay, 1999; Fazio, 1994, 1996). Children with SLI performed lower in rote counting than typically developing children of the same age (Donlan et al., 2007). Furthermore, Cowan, Donlan, Newton and Lloyd (2005) and Donlan et al. (2007) found that difficulties in producing the spoken number sequence, as well as poor comprehension of language, are significantly associated with calculation. Kleemans et al. (2011, 2012) found a relationship between grammatical ability and early numeracy skills. Similarly, neurocognitive studies of adults suggest that linguistic processes such as phonological awareness and grammatical ability are related to both, addition and subtraction (Baldo and Dronkers, 2007; Dehaene et al., 2003). However, the relationship between these skills is complex, and runs counter to other findings which indicate independence between verbal and nonverbal calculation skills (Nunes and Bryant, 1996; Jordan, Huttenlocher and Levine, 1994).

A few recent studies have started to address issues around the association between mathematical language use and the development of mathematical concepts. Saxe, Guberman, Gearhart, Gelman, Massey and Rogoff (1987) reported that mothers from middle socio-economic background engaged their children in more complex math activities and language than low-SES mothers. Therefore, the four year old children from middle-SES outperformed their low-SES peers on complex mathematical tasks. Klibanoff, Levine, Huttenlocher, Hedges, and Vasilyeva (2006) explored the relation between pre-school teachers' amount of math talk and children's growth of mathematical knowledge. They found that the amount of math talk had an effect on the growth of conventional mathematical knowledge. However, Boonen, Kolkman and Kroesbergen (2011) reported not only positive associations with children's number concepts, but also negative associations, possibly indicating confusion where the level of teachers' talk exceeded children's understanding. There is a need for further research concerning the effects of both quantity and quality of math talk on mathematical learning in the early years.

Some studies indicate that the same underlying mechanisms that are crucial for reading attainment may also play an important role in mathematics, particularly phonological awareness (Jordan, Kaplan and Hanich, 2002; Simmons and Singleton, 2008). A recent study by Vukovic and Lesaux (2013) explored the difference between the relationship of children's general verbal abilities and arithmetic compared to phonological skills. The results suggest that eight-year-olds performance on verbal analogies was indirectly related to arithmetic skills through symbolic number skills whereas phonological skills such as phonological decoding were directly related to arithmetic knowledge. The authors concluded that children's general verbal activities may affect children's understanding and reasoning with numbers and children's phonological skills may play a part in executing conventional arithmetic.

In particular, relational terms (e.g. *more, less*) play a crucial role in mature mathematical communication. Although little is known about their acquisition and development, previous studies have shown that young children do not perfectly comprehend relational concepts (Donaldson and Balfour, 1968). Younger children tend to treat *less* as if it is a synonym of *more* conceivably due to the children's bias to choose the greater of two arrays in comparison judgement tasks (Clark, 1973). The implications of these findings for mathematical development have yet to be fully explored. LeFevre and colleagues (2010) assessed four- to seven-year olds vocabulary, phonological awareness, subitizing skills and a spatial span task. They showed that the linguistic, spatial and quantitative pathways contribute independently to the development of early numeracy as well as children's formal mathematical knowledge. However, only vocabulary and phonological awareness were tested to assess linguistic skills, leaving open the possibility that more complex language skills may make further contributions.

1.3 Mathematics and the Approximate Number System.

Numerous studies propose that an innate approximate number sense (ANS; Dehaene, 1992) is fundamental for children's understanding of abstract, symbolic number concepts thus contributing to the development of arithmetic. The ANS is typically assessed using magnitude comparison tasks (comparing numerosities of groups of objects (Barth, Kanwisher, and Spelke, 2003; Piazza, Facetti, Trussardi, Berteletti, Conte, Lucangeli, Dehaene, and Zorzi, 2010)). Performance on these discrimination tasks varies according to the difference between the numerosities, and response time is shorter than is possible by counting. Support for the innate ANS comes from studying infants' discrimination abilities. Xu and Spelke (2000) reported that six-month old infants can successfully discriminate arrays of 8 vs. 16 dots, a ratio of 1:2; by adulthood the discriminability ratio has reduced to 9:10 (Halberda, Mazzocco, Feigenson, 2008).

Evidence in support of the importance of the ANS comes from correlational studies showing that individual differences in ANS and general mathematical achievement were strongly correlated (Halberda et al., 2008; Gilmore, McCarthy, and Spelke, 2010; Libertus, Feigenson and Halberda, 2013), most notably by Mazzocco et al. (2011), who found a significant correlation between ANS precision at pre-school and mathematical skill measured 30 months later. Halberda et al. (2008) found that individual differences in nonsymbolic dot comparison scores in 14-year-old children were correlated with children's past performance on standardised math achievement tasks. Although this study included numerous covariates, the study had an unusual retrospective design. Given that ANS and other cognitive processes develop throughout childhood, the retrospective design makes it problematic to draw a conclusion. Gilmore, McCarthy and Spelke (2007) proposed an alternative approach to the ANS as the developmental precursor of arithmetic. They examined six-year-olds' symbolic approximate arithmetic skills showing that children with no formal instruction on arithmetic can accurately solve approximate arithmetic problems. The authors claim that children's performance on symbolic approximate arithmetic tasks depends on nonsymbolic representations, and provides the basis for exact arithmetic.

In contrast, some studies have failed to report a significant relation between nonsymbolic ANS measures and arithmetic (Holloway and Ansari, 2009; Iuculano, Tang, Hall and Butterworth, 2008; Sasanguie, Göbel, Moll, Smets and Reynvoet, 2012; Kolkman, Kroesberger and Leseman, 2012; Vanbinst, Ghesquière and De Smedt, 2012). Fuhs and McNeil (2013) posited the hypothesis that the link between arithmetic and ANS measures was mediated by children's inhibition skills because performance on nonsymbolic magnitude comparison may rely on children's ability to supress other salient features of the dot arrays such as density or dot size. They assessed pre-school dot comparison skills, mathematics achievement and inhibition abilities in children from low-income background and found that nonsymbolic ANS was a borderline predictor of math achievement. This link was no longer significant once children's inhibition scores were accounted for. The same evidence was found in school children, where children's dot comparison performance did not explain mathematical achievement after performance on inhibition task had been taken into account. However, inhibition was significantly contributing to variance in math scores over and above children's performance on dot comparison (Gilmore, Attridge, Clayton, Cragg, Johnson, Marlow, Simms and Inglis, 2013).

Further evidence from a study by Holloway and Ansari (2009) called the specificity of the correlation between ANS and arithmetic into question. In their study they assessed six- to eight-year-olds performance on a symbolic number comparison task (identifying the greater of two single digit numbers) as well as a nonsymbolic number discrimination task and examined the specific correlations with mathematical achievement. While symbolic and nonsymbolic comparison were highly related, only the symbolic task proved to be predictive of mathematical skills. Subsequent extensive reviews (de Smedt, Noël, Gilmore, and Ansari, 2013; Siegler 2016) have shown that the relationship between the nonsymbolic ANS and children's arithmetic skills is not consistent, while the association between symbolic comparison and mathematic skills is relatively strong. Studies investigating the concurrent correlations between ANS and arithmetic neglect crucial longitudinal aspects of this relationship.

However, recent longitudinal studies produced mixed results (Desoete, Ceulemans, De Weerdt, and Pieters, 2012; Lyons, Price, Vaessen, Blomert, and Ansari, 2014; Kolkman et al., 2012). A longitudinal study by Lyons and colleagues (2014) explored the prediction of arithmetic across grades 1-6 and found no evidence that children's performance on nonsymbolic comparison was a unique predictor of arithmetic scores at any grade. Kolkman et al. (2012) examined the relationship between arithmetic and nonsymbolic, symbolic and mapping skills at age four, five and six. The findings suggest that nonsymbolic, symbolic and mapping skills were separate skills at a younger age integrating over time to form one general numeracy skills concept. Only children's mapping skills were uniquely predictive of math performance at six years.

A recent study by Göbel, Watson, Lervåg and Hulme (2014) addressed the relation between nonsymbolic and symbolic judgement tasks and their role as longitudinal predictors of arithmetic development in six-year-olds. The authors reported that symbolic and nonsymbolic magnitude comparisons define a unitary factor, which was a strong longitudinal correlate of arithmetic skills. The study also included a number identification task in which spoken numerals were presented to be matched to the corresponding Arabic numeral, with a range of targets including single, double and three digit numbers. This measure was not associated with the magnitude comparison factor, and was entered in a longitudinal path model as a separate latent variable, alongside magnitude comparison and other potential predictors of later arithmetic skills, including vocabulary size and nonverbal ability. The path model revealed that number identification was the only significant longitudinal predictor of arithmetic skills a year later, apart from the auto-correlate.

The authors interpret their number identification task as tapping individual differences in both Arabic digit knowledge and place value understanding, suggesting that the former may represent a critical foundational skill underlying early arithmetic, analogous to the role of letter knowledge in reading, and the latter may be crucial for further arithmetic development.

By focussing precisely on properties of the symbol system, these findings offer clarification of previous literature as follows: On the one hand, consistent findings of high correlation between single digit comparison and nonsymbolic comparison (Holloway and Ansari 2009; Göbel et al., 2014; Matejko, and Ansari 2016) reflect general properties of magnitude comparison (as reported by Moyer, and Landauer 1967), relevant but not central to arithmetic development. On the other hand, the specific relation between symbolic comparison and early arithmetic skills (Holloway, and Ansari 2009; Lyons et al. 2014) reflects the contribution of symbolic item identification as a foundational arithmetic skill, but is limited by the nature of the task and the range of single digits. Therefore, when a comprehensive number identification task is included in a longitudinal model of early arithmetic development, simple magnitude comparison fails to predict outcome.

1.4 Mathematics and Cognitive Factors.

Working memory is suggested to play an important role in the development of numeracy skills (Berg, 2008; Geary, Hoard, Byrd-Craven, Nugent, and Numtee, 2007; Jarvis and Gathercole, 2003). Working memory is described as the ability to mentally maintain and manipulate information for a short period of time, storing and accessing information in long-term memory. The dominant model of working memory comprises a phonological loop, visuo-spatial sketchpad, episodic buffer and central executive (Baddeley, 2000).

The majority of research exploring working memory and mathematics achievement has studied primary school children and focused on formal aspects of mathematics. Children with mathematics difficulties tend to perform lower on working memory tasks than their peers (Geary, Hoard, Nugent, and Bailey, 2012; Passolunghi and Siegel, 2004), and there is evidence that working memory is a significant predictor of later mathematics success through middle school (Nunes, Bryant, Barros, and Sylva, 2012).

There is also some evidence that working memory and informal mathematics skills are related in pre-school and kindergarten (Bull et al., 2008; Chiappe, Hasher, and Siegel, 2000). Östergren and Träff (2013) assessed the relation of working memory to informal and formal mathematics skills. Their latent variable model found that verbal working memory was a predictor of both informal and formal skills, and more advanced mathematical concepts. Purpura and Ganley (2014) found that pre-school children's (four- to six-year olds) performance on language tasks was a strong predictor of a range of different mathematical and numeracy competencies, but verbal working memory was only related to cardinality (counting a subset), set comparison and number order.

Some studies reported that the three components of working memory are differentially related to mathematics (Simmons, Willis, and Adams, 2012; Wilson and Swanson, 2001). According to Berg (2008), both phonological loop and visuo-spatial sketchpad function related to arithmetic in children at eight years of age. Third to sixth graders' processing speed, short-term memory, verbal working memory and visual–spatial working memory abilities were assessed. Furthermore, various studies document important relationships between working memory and language and mathematical attainment in school children (Gathercole, Pickering, Knight and Stegmann, 2004; Jarvis and Gathercole, 2003). Gathercole et al. (2004) investigated the relationship between working memory skills including executive functioning (measured by backwards digit span and listening recall) and phonological loop (measured by digit recall and word list matching) and outcome on national curriculum assessments in English, mathematics and science in 7 and 14 year old children. The results noted were that children's performance in English and mathematics was significantly related to working memory skills, and complex span tasks in particular, at 7 years of age. The strong associations between children's performance on complex span tasks and mathematics and science persisted at 14 years, but the link between working memory and English was not significant suggesting that cognitive processes required in the curriculum areas of mathematics and science may be influenced by general capacities of working memory.

Additionally, central executive functioning may relate to children's emergent mathematical achievement (Bull, Espy, Wiebe, Sheffield and Nelson, 2011). Bull et al. (2011) assessed pre-school children's (four years) performance on a comprehensive battery of executive functioning tasks, as well as age and vocabulary skills. Using SEM models, the authors report that age, sex and social factors affect mathematical achievement. Age differences contribute indirectly to emergent mathematical skills via central executive functioning. The pattern that emerged was that older children show a better developed central executive functioning than their younger peers and this age-central executive relation is significantly higher in girls than boys, which may point to sex differences in the development of the frontal lobe. However, the study did not assess other possible precursors of arithmetic.

Executive functioning can be divided into three types: monitoring and manipulating information (working memory), supressing unwanted and distracting information (inhibition) and flexible thinking (shifting) (Gilmore, Keeble, Richardson and Clagg, 2014). A growing body of research has turned its focus on the link between inhibition and mathematical attainment. Most studies examined school aged children (Gilmore et al., 2014; Visu-Petra, Cheie, Benga and Miclea, 2011; St Clair-Thompson and Gathercole, 2006). It has been reported that children's performance on inhibition tasks is related to their school outcome in mathematics (Brock, Rimm-Kaufman, Nathanson and Grimm, 2009; Visu-Petra et al., 2011) and children's performance on standardised mathematical tests (St Clair-Thompson and Gathercole, 2006). Wang, Tasi and Yang (2012) reported that children's performance on inhibition tasks was poorer for children with mathematical learning difficulties compared to their normally developing peers. Gilmore et al. (2014) were exploring performance on inhibition tasks and mathematical achievement, and factual, procedural and conceptual arithmetic knowledge in particular, in older children. They found that inhibition scores were strongly associated with arithmetic and that inhibition influenced conceptual knowledge in older children and procedural skills in younger children.

In contrast, many studies have failed to find a link between children's inhibition performance and arithmetic skills. Waber, Gerber, Turcios, Wagner and Forbes (2006) showed a weak link between inhibition skills and mathematical performance. A few studies reported that inhibition skills were not a unique predictor of mathematical performance (Miller, Müller, Giesbrecht, Carpendale and Kerns, 2013; Monette, Bigras and Guay, 2011). Similar, there is evidence that the link between inhibition skills and mathematical performance may no longer be significant once children's shifting abilities are accounted for (Bull and Scerif, 2001; Van der Ven, Kroesbergen, Boom and Leseman, 2012).

There is mixed evidence concerning the link between inhibition and arithmetic. Interestingly, Bull et al. (2011) analysed pre-school children's arithmetic skills using SEM models and found a significant prediction of arithmetic by executive functioning. However, the study did not control for other covariates of arithmetic. Little is known of the importance of cognitive components in very early numeracy skills. Further research is therefore needed to expand the existing evidence and add to the knowledge of the fundamental underlying skills that contribute to early arithmetic and how the contribution of precursors change over time.

1.5 Early arithmetic and number knowledge.

While there is evidence of the developmental importance of ANS for arithmetic (and the possibility that ANS drives exact arithmetic via approximate arithmetic skills), and also for the role of general cognitive abilities (executive functioning in particular) as drivers of emergent arithmetic skills, the possibility remains that number knowledge itself is the major precursor. Jordan et al. (2009) investigated the relation between early number competence and mathematics achievement from beginning of kindergarten to the middle of grade 1. They showed that kindergarten number competence predicted rate of growth in mathematics achievement. However, their number competency factor comprised of a wide range of numerical skills including counting and number recognition, number comparisons, nonverbal calculation, story problems, and number combinations.

Göbel et al. (2014) noted that number identification, in which spoken numerals were presented to be matched to the corresponding Arabic numeral (targets included single, double and three digit numbers), was the most powerful longitudinal predictor of arithmetic skills a year later, apart from the auto-correlate. Magnitude comparison, though highly correlated with later arithmetic, was not a significant predictor once number identification was taken into account. These results are more suggestive of the particular importance of children's associations between spoken and written symbols. The findings show the importance of early number knowledge for driving children's mathematical learning in school.

1.6 Purpose of this Thesis.

Recent research (Lyons et al. 2014) highlights the time-sensitive nature of influences on mathematical development. Might it be the case, then, that the findings of Göbel at al. (2014) represent a transient state in which symbolic knowledge has particular importance? Most importantly, if similar measurements were taken during preschool would they support the proposal that knowledge of the symbol system drives later arithmetic development, or would they support the findings of Mazzocco et al. (2011) indicating that early precision in nonsymbolic magnitude comparison, measured before the onset of formal schooling, provides the basis for later arithmetic skills, or would they attribute greater importance to language or general cognitive abilities?

The purpose of this thesis was to examine a broad range of possible longitudinal precursors of early arithmetic skills over a 25-month period using structural equation modelling. The comprehensive test battery explored children's early arithmetic performance and included assessment of numeracy, language and cognitive abilities. Based on the premise that mathematical concepts are fundamentally underpinned by linguistic representations, the thesis aims to clarify the association between arithmetic skills and language comprehension, and to explore the possible importance of specific mathematical language. Furthermore, the study investigated to what extent ANS on the one hand and specific number knowledge place critical constraints on the development on early arithmetic.

Children at age four attending preschool (morning or afternoon sessions only) were assessed on their general and specifically mathematical language skills, and on a range of symbolic and nonsymbolic magnitude comparison tasks, adjusted to capture earlier developmental levels. An expanded set of tasks to capture number knowledge, including reading and writing numbers as well as number identification, were administered: following Mix, Prather, Smith and Stockton (2014) a wide range of multi-digit numbers were included. Rote counting ability, vocabulary and grammatical comprehension and general cognitive ability (nonverbal intelligence) were controlled for.

The thesis asks the following questions. What is the extent of preschool children's knowledge of the Arabic number system? Is Arabic number knowledge closely associated with counting ability? Do symbolic and nonsymbolic magnitude comparison tasks form a unitary factor (as in Göbel at al., 2014)? What is the importance of specific mathematically related language? Which of the latent variables formed by our predictor variables account for arithmetic skills at age six (after a year of formal schooling)?

Furthermore, the study aimed to establish a detailed model of the typical development of numeracy, thereby facilitating subsequent research into the development of a screening tool to identify children at risk of numeracy difficulties.

The main goals of this thesis are therefore:

- To explore the relationship between symbolic and nonsymbolic magnitude comparison tasks and whether they are distinct or rely on the same underlying construct.
- To assess the performance and structure of the magnitude comparison tasks over a 25-months period, within the same sample of children, in order to capture developmental change.
- To assess the importance of specific mathematical language.
- To assess the importance of number knowledge.
- To investigate the concurrent predictors of early arithmetic in a large represeentative sample.
- To assess longitudinal relations between ANS, numerical knowledge, language and general cognitive abilities and arithmetic in a large representative sample at a critical period of math development (transition from pre-school to formal schooling) over a 25-months time period.
- To assess whether the link between ANS and math achievement was mediated by children's inhibition skill.
- To explore the nature and development and relationship of symbolic and nonsymbolic approximate arithmetic.
- To investigate pre-schoolers‘ performance on number line estimation tasks and its contribution towards early arithmetic.

Chapter 2. Methods

A longitudinal, multifactorial study was carried out looking at predictors of children's early arithmetic skills. Structural equation modelling was used to examine the relationship between language, ANS and early arithmetic in typically developing children. Background variables also included intelligence, memory components and various language as well as number knowledge tasks examining different aspects of numerical understanding and processing. The analyses took account of any statistical constraints occurring in the data. Data were collected through a comprehensive test battery on cognition, language and numeracy skills.

2.1 Participants.

Typically developing children in one UK public primary school (Chafford Hundred, Essex) were assessed five times over a 25-month period from summer term of nursery (age four) to the summer term of Year One of formal schooling (age six). To establish socioeconomic status, participants were assigned to the Office for National Statistics' lower super output areas (LSOAs) based on the school's postcode. LSOAs are new geographical areas for reporting small-area statistics in England (Neighbourhood Statistics, 2016, Look ups, 2016). It can be assumed that most children live either in the same or adjacent areas. The Index of Multiple Deprivation (IMD) 2015 was computed based on the LSOA using Neighbourhood Statistics (2016). The 2015 indices of deprivation consisted of seven dimensions (income, employment, health, education, barriers to housing and services, crime and living environment) which were combined to the overall IMD. The most deprived neighbourhood in England has the rank one and the least deprived has the rank 32,844.The overall IMD of the LSOA Thurrock 020D indicates that this area shows lesser than average deprivation: 8th percentile (where 1 is most deprived 10%), rank 25,631 out of 32,844. This is comparable to a middle socio-economic background.

The specific means and standard deviations of the children's age at each time of testing are displayed in Table 2.1. Participation was voluntary and children were informed prior to testing that they could withdraw at any time.

Table 2.1

Mean age and standard deviations of children

	Time 1		Time 2		Time 3		Time 4		Time 5	
	Summer term of nursery		Spring term of reception		Autumn term of Year One		Spring term of Year One		Summer term of Year One	
Gender	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*
Boys	50.87 (3.51)	48	59.21 (3.38)	62	66.36 (3.27)	62	70.28 (3.29)	61	75.83 (3.34)	66
Girls	4.50.34 (3.56)	52	58.83 (3.60)	55	66.10 (3.63)	54	70.04 (3.63)	54	75.69 (3.65)	53
Total	50.60 (3.53)	100	59.03 (3.47)	117	66.24 (3.43)	116	70.17 (3.44)	115	75.76 (3.46)	119

Notes. M = mean age in months. SD = standard deviation in months. Time 1 was assessed in May-June 2014, Time 2 in February-March 2015, Time 3 in September 2015, Time 4 in January 2016 and Time 5 in June 2016.

2.2 Materials.

At Time 1, most background measures as well as language and numeracy measures were administered whereas only language and numeracy measures were reassessed at Time 2 to Time. Each task will be explained in full detail in the relevant results chapters.

2.2.1 Measures taken at Time 1.

All tasks were administered individually to the four nursery classes in the summer term of the nursery age (4 years). The following measures were taken at Time 1: Nonverbal intelligence (Raven's Coloured Progressive Matrices (Raven's CPM Raven, Court, and Raven, (1993)), central executive functioning, phonological loop, grammatical ability (Test for Reception of Grammar II (TROG-2 Bishop, 2003)), vocabulary (British Picture Vocabulary Scale 3rd Edition (BPVS - III Dunn, Dunn and Styles, 2010)), reading skills and letter writing skills, specific math-related language ability, number knowledge (including number identification, number reading, number writing, rote counting as well as numerical estimation (number line task)), measures of approximate number system (magnitude comparison) and arithmetic skills (simple addition).

2.2.2 Measures taken at Time 2.

All tasks were administered individually. Due to the fact that many tasks showed ceiling effects at Time 1, difficulty levels were adjusted accordingly, taking children's age and experience with the task into account. Only literacy (reading and letter writing skills), math-related language, magnitude comparison, number knowledge and arithmetic were re-assessed at Time 2 in the spring term of the reception year (9 months later).

2.2.3 Measures taken at Time 3.

Most Time 2 tasks, as well as number line estimation were re-assessed at Time 3 in the autumn term of Year One (16 months after Time 1). Two new arithmetic tasks and one behavioural regulation task were introduced to the test battery, now that the children entered formal schooling. Similar to Time 2, difficulty levels were adjusted where necessary.

Tasks were re-designed as group tasks, where possible, to shorten testing time and hence reducing the cognitive load on children. The rationale behind it was that children in Year One are able to handle the group setting based on the structure of their regular school day and groups were kept small with no more than five children being tested at a time by at least two experimenters. Moreover, only already familiar tasks were introduced as group tasks. However, not all tasks could be re-designed as group tasks. The test battery comprised measures of central executive functioning, behavioural regulation (Head-to-Toes task), math-related language ability, number knowledge (including number line task), magnitude comparison and arithmetic (Test of Basic Arithmetic and Numeracy Skills (TOBANS Brigstocke, Moll and Hulme, 2016)) and approximate arithmetic (Gilmore et al., 2007)).

2.2.4 Measures taken at Time 4.

Most Time 3 tasks were re-assessed at Time 4 in the spring term of Year One (20 months after Time 1). Difficulty levels were adjusted where necessary. Tasks included: central executive functioning, behavioural regulation, number knowledge (except for counting and number line task), magnitude comparison and both measures of arithmetic (TOBANS and approximate arithmetic).

2.2.5 Measures taken at Time 5.

Most Time 3 and Time 4 tasks were re-assessed in the summer term of Year One (25 months after Time 1), except for counting and the inhibition task Head-Toes-Shoulders-Knees-task. Two new literacy tasks and one arithmetic task were introduced to complete the test battery. Difficulty levels were adjusted where necessary. The test battery included measures of central executive functioning, reading (PWM task used in Time 1 and Time 2 as well as the new measure Test of Word Reading Efficiency–Second Edition (TOWRE–2; Wagner, Torgesen and Rashotte, 2011)) and spelling skills (Single Word Spelling Test (SWST; Sacre and Masterson, 2000)), number knowledge tasks (including number line task), magnitude comparison as well as arithmetic (TOBANS, approximate arithmetic and Numerical Operations subtest of the second edition of the Wechsler Individual Achievement Test (WIAT-II; Wechsler, 2005)).

2.3 Procedure.

All tests were administered in a separate room or another quiet place in the school. Group testing (employed at later ages) was used in a small group setting with not more than five children at Time 4 and not more than eight at Time 5. Tests were divided into counterbalanced sessions of 20 to 40 minutes (4 sessions at Time 1, 3 sessions at Time 2, 2 sessions at Times 3, 4 and 5). Testing was carried out five times over a 25-month period from the summer term of nursery (May-June 2014) through to the summer term of Year One (further testing sessions took place in February-March 2015, September 2015, January 2016 and June 2016). Not all tests were re-administered at all the time points.

Wherever possible, each child was seen by the same experimenter. The author was assisted by several research assistants from undergraduate psychology classes. They were trained on how to administer the test battery; all were experienced in working with young children. Children were tested individually at Times 1 and 2, and at Times 3, 4 and 5 individually or in small groups in a separate room or another quiet place in the school. Each child met with the experimenter ideally two to four days in a row, depending on the number of blocks, to enhance motivation or concentration. If testing in groups, the ratio of experimenters to children was 1:3.

Preliminary to the testing, the experimenters attended at least one day in each class so that the children got to know them and felt more comfortable around them. Moreover, the experimenters told each child that they would play games and asked questions such as “How are you?” or “How old are you?” prior to each testing session.

All unstandardized tests included practice items. Concerning feedback, children received only concrete feedback on their performance for practice items and general praise and encouragement throughout the tests.

2.4 Data Analysis.

Descriptive statistics as well as simple analyses of variance (ANOVA), t-tests and simple regressions were conducted using IBM SPSS Statistics 22. Significance alpha level was chosen prior to be .05.

2.4.1 Structural Equation Modelling (SEM).

The purpose of this study was to examine precursors of early numeracy skills and their relation to language and magnitude comparison in particular. The association between basic arithmetic skills and early cognitive skills, language comprehension and numeracy skills was assessed, thus determining the extent to which these precursors and arithmetic are related. The study therefore addressed to what extent early arithmetic skills can be predicted from cognitive and linguistic measures.

Structural Equation Modelling (SEM) is a powerful statistical analysis tool for comprehensive models. Its origins are in ‘regression analyses of observed variables and in factor analyses of latent variables’ (Kline, 2016, p.24). In SEM, observed (manifest or measured data) variables are distinguished from latent variables. Those latent variables correspond to hypothetical constructs which are not directly observable. SEM first tests whether a prior hypothesised model fits the measured data and then whether the relationships between these latent variables are significant. This unique blend of regression and factor analysis is the reason why it is the chosen statistical tool for the main data analysis of this thesis.

As regression analyses are part of SEM, the common assumptions for regression are also true for SEM.

1. Linear relationship between regression coefficients (including manifest and latnet variables aas well as among latent variables).
2. Normality and homoscedasticity of residuals.
3. No or little multicollinearity.
4. No outliers or missing data (however, Mplus Version 7 (Muthén and Muthén, 2013) deals with missing data using the maximum likelihood method).
5. Large enough sample size.

The commonly used rule of thumb for SEM is the ratio 10:1 (observation data/latent variable), for example you need to test 10 children per each latent variable. The most complex model in this thesis comprises of 9 latent variables and hence a sample size of 90 children should be adequate. However, few simulations studies showed that there is no one-fits-all solution to sample size recommending rather small sample sizes as enough. Wolf, Harrington, Clark and Miller (2013 found that sample size requirements ranged from 30 subjects for simple confirmatory factor analysis with four indicators and loadings around .80 and up to 450 subjects for mediation models. Similarly, Sideridis, Simos, Papanicolaou and Fletcher (2014) noted that a sample size of 50-70 subjects would be enough for a model of functional brain connectivity involving four latent variables.

With a large enough sample size SEM techniques such as confirmatory factor analysis (CFA) or path analysis can be employed. CFA is commonly used to test whether, and to what extent, measured data (manifest variables) are underpinned by the same constructs and processes by forming latent variables (factors) of the underlying construct. Therefore, CFA tests whether the data fit a hypothesized factor structure which eventually can be applied to run path models. Moreover, SEM also gauge the fit of these proposed models to the observed data. CFAs can form the basis of path models that estimate the prediction of children’s early arithmetic skills.

Variables presented in rectangles reflect manifest variables (observed data), while ellipses represent latent variables (hypothesised factors) that form an underlying construct. One-headed arrows from the latent factor to the manifest variables depict causal paths and residuals of each construct (unexplained variance of the measure) are reflected by one-headed arrows pointing towards the latent variable. Values ascribed to these connections are standardised regression coefficients (factor loadings) thus it is possible to compare the values. Two-headed arrows reflect true-score correlations between constructs. Solid lines illustrate statistically significant relationships, and dashed lines illustrate statistically nonsignificant relationships. Missing data were estimated using MPlus' default mode of maximum likelihood (ML).

To assess the goodness of fit between sample data and proposed model, four statistics will be reported:

1. Chi-square. The chi-square difference test assesses how the covariance matrix of the proposed model deviates from the covariance matrix of the sample (Byrne, 2012). Non-significant chi-square indicates that two models do not differ. However, chi-square difference test is highly sensitive to sample size and a large sample size may cause a significant result.

2. Root Mean Square Error of Approximation (RMSEA). This index of goodness of fit estimates the fit of the hypothesized model. It has been recommended that a value of a well-fitting model is less than .06 (Hu and Bentler, 1999) or less than .05 (Browne and Cudeck, 1992).

3. Comparative Fit Index (CFI). Ranging from 0 to 1, an ideal value is close to one. Hu and Bentler (1999) suggest a value of greater than .95 for a good fit.

4. Standardized Root Mean Residuals (SRMR). This is the final reported index of model fitness. Byrne (2012) recommends a value less than .05 for a well-fitting model.

Chapter 3. Development of Magnitude Comparison: Moving from a two-factorial model towards a unitary model

A primary aim of the study was to assess the underlying factor structure of multiple measures of the ANS and to maximize the reliability of the measurements. This was addressed by investigating whether different measures of magnitude comparison cohere to define multiple or a single construct and whether this structure changes over time.

Numerous studies propose that ANS is fundamental for children's understanding of abstract, symbolic number concepts thus contributing to the development of arithmetic. Evidence in support of the importance of the ANS comes from correlational studies showing that individual differences in nonsymbolic ANS and general mathematical achievement were strongly correlated (Halberda et al., 2008; Gilmore et al. 2010; Libertus et al., 2011; Mazzocco et al., 2011).

In contrast, some studies have failed to report a significant relation between nonsymbolic ANS measures and arithmetic (Holloway and Ansari, 2009; Iuculano et al., 2008; Sasanguie et al., 2012, Kolkman et al., 2013; Vanbinst et al., 2012). Holloway and Ansari (2009) assessed six- to eight-year-olds performance on a symbolic number comparison task as well as a nonsymbolic number discrimination task and examined the specific correlations with mathematical achievement and found that only the symbolic task proved to be predictive of mathematical skills. Subsequent extensive reviews (de Smedt et al., 2013; Siegler 2016) have shown that the relationship between the nonsymbolic ANS and children's arithmetic skills was not coherently evident, while the association between symbolic comparison and mathematic skills is relatively strong.

Recent longitudinal studies produced mixed results (Desoete et al., 2012; Lyons et al., 2014; Kolkman et al., 2012). A longitudinal study by Lyons and colleagues (2014) explored the prediction of arithmetic across grades 1-6 and found no evidence that children's individual differences on nonsymbolic comparison was a unique predictor of arithmetic scores at any grade.

Few studies explored the internal factorial structure between the various symbolic and nonsymbolic magnitude comparison tasks. Kolkman et al. (2012) examined the relationship between arithmetic and nonsymbolic, symbolic and number line skills at age four, five and six. The findings suggest that nonsymbolic, symbolic and number line skills were separate skills at a younger age integrating over time to form one general numeracy skills concept. Only children's number estimation skills were uniquely predictive of math performance at six years. Similarly, a recent study by Göbel et al. (2014) addressed the relation between nonsymbolic and symbolic judgement tasks and their role as longitudinal predictors of arithmetic development in six-year-olds. The authors reported that symbolic and nonsymbolic magnitude comparison tasks were best described as one general magnitude comparison factor. The path model revealed that number identification was the most powerful longitudinal predictor of arithmetic skills at age seven, apart from the auto-correlate.

To my knowledge no study has fully investigated commonalities and differences between symbolic and nonsymbolic tasks and whether their relationship changes over the time period of pre-school to early formal schooling. To address this research question, various symbolic and nonsymbolic magnitude comparison tasks were created as widely used in the literature (e.g. Göbel et al., 2014; Holloway, and Ansari, 2009; Kolkmann et al., 2013). Symbolic and nonsymbolic distance effects were assessed using the same number pairs as either Arabic numerals or arrays of squares. Ratios for nonsymbolic comparison were chosen based on Halberda and Feigenson's (2008) finding that four and five year olds can correctly solve more difficult ratios such as 2:3, 3:4 or 5:6.

The aim of this chapter was to assess the underlying latent factors of symbolic and nonsymbolic magnitude comparison tasks. Do the two represent one general magnitude comparison latent factor or two distinct latent factors? And does this underlying structure change over time?

3. 1 Methods.

3.1.1 Participants.

The same participants were used as described in Chapter 2 (p. 42)

3.1.2 Materials.

3.1.2.1 Times 1 and 2. Various comparison tasks were created for the study. Each comparison pair was presented on a single page (Appendices 1). Children were given one point for every correct comparison with a maximum score of 16 for each comparison subtask and 160 overall.

Symbolic Digit Comparison Task. Pairs of Arabic numerals were displayed within two adjacent boxes (12cm x 12cm) with digits in Calibri font size 350. Digits ranged from one to nine and both orders of the pairs were presented (e.g. 3 and 4 versus 4 and 3). To investigate the numerical distance effect (Moyer and Landauer, 1967), two versions were administered. In the close version, the difference between the two digits was one or two and in the far version, the difference was five, six or seven.

Nonsymbolic Magnitude Comparison Tasks. Nonsymbolic comparison tasks consisted of arrays of black squares within 12cm x 12cm boxes similar to the symbolic task. In the fixed size condition, numerosities presented were ranging between 5 and 13. In the close version of the fixed size condition the difference between arrays was one or two squares and in the far version of the fixed size condition the difference between arrays was five, six or seven.

The size of the squares included in the arrays was manipulated as follows: In the fixed size condition, all squares were of the same size. In the surface area matched condition, the size of the squares was controlled so that total surface area was matched across arrays within stimulus pair, so that smaller numerosities had larger squares and larger numerosities had smaller squares.

In the surface-area matched condition, larger numerosities ranging from 20 to 40 were examined. Similar to Göbel et al. (2014), baseline numbers 20 through to 30 were compared to their nearest whole number in the ratios 2:3 and 3:4, thus 23 was compared to 35 (2:3) or 31 (3:4).

3.1.2.2 Times 3, 4 and 5. A recent study (Göbel et al., 2014) showed that children in Year One can successfully perform magnitude comparison tasks in a group setting. Thus the magnitude comparison task used in this study was redesigned as a group test using the same stimuli pairs created at Times 1 and 2. Symbolic and nonsymbolic comparisons were presented in pairs of two adjacent 2.1 cm x 2.1 cm boxes. The same 12 variations of size, ratio and distance effect as in Times 1 and 2 were presented in two booklets matched for difficulty level. The order of target locations (left array vs. right array) was controlled in order to avoid repeated response patterns. Each variation comprised of 36 item pairs. Six of the pairs were displayed on each page and there were six pages per subtask (Appendix 19 for order of subtasks).

Children were asked to tick the bigger number or box with more dots. Two practice trials were displayed on the first page of each subtask. The first trial was demonstrated by the experimenter who then asked children to tick the next box. Another six practice items were then given to the children, but only for the first three subtasks of the first booklet. Feedback was given on practice items but not on test trials. Children had 30 seconds per subtask to solve as many comparisons as possible. The order of the two booklets was counterbalanced, with half of the children starting with booklet A and the other half with booklet B.

3.1.3 Procedure.

The magnitude comparison tasks were assessed as part of a larger test battery. Children were told not to count the dots, but choose the bigger array as quickly as possible. If a child attempted to count the dots, the experimenter reminded the child not to count the dots. At Times 1 and 2, magnitude comparison tasks were individually administered, split up into three (Time 1) or two (Time 2) parts, in a separate room or quiet place in the school. To discourage counting, the experimenter displayed the pairs of stimuli for a short time and encouraged the child to choose the right stimulus as quickly as possible.

At Times 3, 4 and 5, the magnitude comparison task was administered as a group task to shorten testing time. The rationale behind this was that children in Year One are able to handle the group setting based on the structure of their regular school day. Groups were kept small with no more than five to eight children being tested at a time (the ratio of experimenters to children was 1:3).

Concerning feedback, children received only concrete feedback on their performance for practice items and general praise and encouragement throughout the tests.

3.2 Results.

The main goal was to identify the factorial structure and reliability of measures of the ANS system by investigating whether different measures of the magnitude comparison cohere to define multiple or a single construct. First, a descriptive analysis of the magnitude comparison tasks was conducted followed by the analysis of distance and ratio effects. Descriptive statistics and simple analyses of variance (ANOVA) were conducted using IBM SPSS Statistics 22. To answer the research question concerning the development of the ANS at pre-school age, a series of confirmatory factor analyses were conducted to investigate the relation between the various magnitude comparison tasks using Mplus Version 7 (Muthén and Muthén, 2013). First, the single factor model (all subtasks comprise one factor) was compared to the two factor model (symbolic versus nonsymbolic comparison tasks). Lastly, the CFAs were re-run using a specific subgroup of the sample (45 children). The subgroup (number wizards; high achievers in number reading) comprised of all children that scored the maximum in the number reading task at Time 1.

3.2.1 Descriptive Statistics.

Descriptive statistics for the different magnitude comparison tasks can be seen in Table 3.1. Children's performance consistently improved over time across all subtasks. Few children scored at ceiling level at Time 1, but clear ceiling effects can be found at most subtasks at Time 2. After introducing the time-constrained version of the tasks at Time 3, no child scored the maximum, and only very few reached the maximum score at Time 5.

Comparing the individual subtasks, children performed less accurate on both symbolic and nonsymbolic close (digit close, fixed size close and surface-area matched close) trials compared to the far trials, supporting the classic distance effect. Similarly, children scored better at the 2:3 ratio (fixed size as well as surface-area matched) and the ratio 5:6 was the most difficult ratio.

3.2.2 Distance Effects.

The distance or ratio effects in comparison tasks are characterised by the finding that it is harder to discriminate between two arrays of items or Arabic numerals that are numerically close than it is to compare stimuli that are numerically distant. According to Weber's Law, accuracy on magnitude comparison tasks increases (and response time decreases) as the numerical ratio increases (results are shown in Table 3.2).

3.2.2.1 Symbolic Distance Effect.

Analyses of variance assessed distance effects across all time points. Two ANOVAs were conducted. The first 2 (distance) x 2 (time) ANOVA examined distance effects between symbolic close and far trials at Times 1 and 2 which were administered individually whereas the second 2 (distance) x 3 (time) ANOVA compared symbolic close and far trials at Times 3, 4 and 5 which were time-constrained group tests. The first ANOVA found significant main effects for distance ($F(1,74) = 100.920$, $p < .001$, $\eta_p^2 = .187$) and time ($F(1,74) = 29.104$, $p < .001$, $\eta_p^2 = .282$). Children performed better on digits far trials ($M = 12.09$, $SD = .31$) compared to digits close ($M = 10.93$, $SD = .31$).

Table 3.1

Mean and standard deviations of predictor and criterion measures from all testing sessions

		Time 1	Time 2	Time 3	Time 4	Time 5
		M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Magnitude	Digit Close	10.11 (3.21) [7]	11.97 (3.07) [20]	10.90 (4.92)	13.04 (4.46)	16.89 (4.57)
Comparison	Digit Far	10.73 (3.59) [7]	13.64 (3.33) [60]	14.32 (4.85)	17.20 (5.33)	22.32 (5.82) [2]
	NS FS Close	10.26 (2.16) [1]	10.64 (2.20) [1]	9.99 (4.14)	12.12 (4.07)	15.14 (4.33)
	NS FS Far	12.87 (2.57) [17]	14.13 (2.04) [34]	15.49 (6.11)	17.99 (5.31)	21.68 (6.32) [3]
	NS FS 2:3		13.18 (2.35) [24]	13.90 (5.12)	17.63 (5.98)	21.98 (6.41) [4]
	NS FS 3:4	10.76 (2.72) [4]	11.97 (2.44) [16]	12.82 (5.38)	15.56 (5.23)	19.68 (5.86) [2]
	NS FS 5:6	10.58 (2.29) [3]	11.23 (2.28) [3]	9.86 (4.95)	12.40 (3.94)	14.81 (4.95)
	NS SA Close	10.29 (2.28) [1]	10.67 (2.00)	8.99 (3.64)	10.57 (3.77)	12.47 (3.93)
	NS SA Far	13.24 (2.43) [23]	13.49 (2.09) [19]	14.48 (5.93)	17.78 (5.45)	21.96 (6.30) [2]
	NS SA 2:3	11.42 (2.42) [5]	12.37 (2.28) [8]	12.68 (5.51)	16.14 (5.42)	19.12 (6.45)
	NS SA 3:4	10.81 (2.03) [2]*	11.56 (2.23) [3]	10.92 (5.14)	13.40 (4.90)	17.35 (5.45)
	NS SA 5:6		10.45 (2.11) [1]*	9.60 (4.13)	11.34 (3.73)	13.23 (4.60)

Notes. M = mean age. SD = standard deviation * individually administered tasks. The number of children scoring the maximum score are shown in square brackets. All scores are presented as raw scores. For the Magnitude Comparison Tasks: NS = nonsymbolic. FS = fixed size trials. SA = surface-area matched trials

Table 3.2

Numerical Distance Effects across Time

				Digits Close	Digits Far
	t value	Degrees of Freedom	p value	M (SD)	M (SD)
Time 1	-1.735	1, 99	.086	10.11 (.321)	10.73 (.36)
Time 2	-6.401	1, 116	<.001	11.97 (.28) <	13.64 (.31)
Time 3	-9.454	1, 115	<.001	9.83 (.44) <	13.25 (.51)
Time 4	-9818	1, 110	<.001	12.98 (.43) <	17.03 (.51)
Time 5	-9.897	1, 116	<.001	15.70 (.46) <	20.02 (.59)

Notes. M = mean. SD = standard deviation. Paired samples t-Tests. Times 1 and 2 were individually administered compared to the time-limited Times 3, 4 and 5 group test.

Also, performance improved over time with children scoring higher at Time 2 (M = 12.55, SD = .34) than Time 1 (M = 10.47, SD = .34). Furthermore, there was a significant interaction ($F(1,74) = 4.504$, $p = .037$, $\eta_p^2 = .057$) between distance and time. Post-hoc t-tests revealed that children performed significantly better on far trials than close trials at Time 2, $t(1,116) = 6.401$, $p < .001$ but the distance effect for the symbolic task at Time 1 (numeric distance effect) was not significant, $t(1,99) = 1.735$, $p = .086$. This may be due to the fact that some children had difficulties reading the Arabic numerals. It was noted that a third of the children made at least two mistakes in reading single digit Arabic numerals (assessed as part of the number knowledge battery). Taking this into account, an ANOVA testing the numeric distance effect in children with a complete understanding of Arabic numerals (defined as all children scoring the maximum score on the number reading task), a marginally significant numerical distance effect was found ($F(1,44) = 3.914$, $p = .05$, $\eta_p^2 = .08$) with children performing better at digits far trials (M = 12.24, SD = .51) than digits close (M = 11.20, SD = .50).

The second ANOVA concerning Times 3, 4 and 5 revealed a significant main effect for distance ($F(1,105) = 212.858$, $p < .001$, $\eta_p^2 = .670$), with children performing more accurately on digits far trials (*M* = 16.97, *SD* = .46) than digits close (*M* = 13.00, *SD* = .39). The main effect for time was also significant ($F(2,210) = 134.279$, $p < .001$, $\eta_p^2 = .561$). Accuracy increased over time. Children performed significantly better at Time 5 (*M* = 18.16, *SD* = .49) compared to Time 4 (*M* = 15.07, *SD* = .43) and Time 3 (*M* = 11.73, *SD* = .46). The interaction was not significant ($F(2,210) = 1.228$, $p = .295$, $\eta_p^2 = .012$).

3.2.2.2 Nonsymbolic Distance Effect. Investigating distance effects for nonsymbolic close and far trials for fixed size and surface-area matched size comparison tasks, a 2 (distance; close versus far) x 2 (size; fixed size versus surface-area matched) repeated measures ANOVAs were run at each time point. The analyses showed clear distance effects across all time points with better performance on trials where the difference between the two arrays of squares was greater.

At Time 1, we found main effects for distance ($F(1,99) = 207.54$, $p <.001$, $\eta_p^2 = .677$), with children performing better on trials where the difference between the squares was large (*M* = 13.06, *SD* = .22) than trials with a smaller difference (*M* = 10.28, *SD* = .17). However, the main effect for size and the interaction were not significant ($F(1,99) = .893$, $p = .35$ and $F(1,99) = 1.031$, $p = .31$ respectively).

At Time 2, the results showed main effects for distance ($F(1,116) = 345.499$, $p < .001$, $\eta_p^2 = .749$), with children performing better on nonsymbolic far trials (*M* = 13.81, *SD* = .17) than nonsymbolic close trials (*M* = 10.65, *SD* = .17), and for size ($F(1,116) = 4.306$, $p = .04$, $\eta_p^2 = .036$), with children performing better on fixed size trials (*M* = 12.39, *SD* = .16) than surface-area matched trials (*M* = 12.08, *SD* = .15). Furthermore, the interaction was also significant ($F(1,116) = 4.872$, $p = .029$, $\eta_p^2 = .04$). Post-hoc t-tests confirmed that the size affects far and close trials differently, with far trials being more affected by size ($t(1,116) = 3.486$, $p = .001$) than close trials ($t(1,116) = -.109$, $p = .91$). Nonetheless, the effect size for both - main effect and the interaction - were small suggesting that these differences may be relatively unimportant.

Comparable to Time 2, the main effect for distance at Time 3 was significant ($F(1,115) = 165.504$, $p <.001$, $\eta_p^2 = .59$). Children performed better on nonsymbolic far trials (*M* = 13.72, *SD* = .52) than nonsymbolic close trials (*M* = 9.06, *SD* = .34). Moreover, the main effect for size was also significant ($F(1,115) = 16.42$, $p <.001$, $\eta_p^2 = .125$). Fixed size trials were easier (*M* = 11.87, *SD* = .42) than surface-area matched trials (*M* = 10.91, *SD* = .42). However, the interaction was not significant (*F*(1,115) = .487, *p* = .49).

Similar to Time 3, both main effects at Time 4 were significant (distance: *F*(1,110) = 497.727, *p* <.001, η_p^2 = .819 and size: *F*(1,110) = 14.058, *p* <.001, η_p^2 = .113) but the interaction was not significant (*F*(1,110) = 2.598, *p* = .110). These results show that children performed better on far trials (*M* = 17.77, *SD* = .45) than close trials (*M* = 11.31, *SD* = .32) and better on fixed size trials (*M* = 14.99, *SD* = .38) than surface-area matched (*M* = 14.08, *SD* = .39).

Analysis of Time 5 showed main effects for distance (*F*(1,116) = 462.274, *p* <.001, η_p^2 = .799) and size (*F*(1,116) = 14.562, *p* <.001, η_p^2 = .112), and a significant interaction (*F*(1,116) = 5.193, *p* = .025, η_p^2 = .043). Children showed greater performance on far trials (*M* = 20.14, *SD* = .56) than close (*M* = 12.48, *SD* = .36), and fixed size trials (*M* = 16.82, *SD* = .47) than surface-area matched (*M* = 15.80, *SD* = .44). Post-hoc t-tests suggest that size affects far and close trials differently, with close trials being more affected by size (*t*(1,116) = 4.870, *p* <.001) than far trials (*t*(1,116) = .762, *p* = .45).

3.2.3 Ratio effects.

Prior to the main testing, a pilot test of ratio effects was carried out on reception class children (one year older than the study sample) which revealed that reception class children struggled with the ratios fixed size 2:3 and surface-area matched 5:6. Hence the targeted nursery cohort was not assessed on those trials but only fixed size 3:4 and 5:3 and surface-area matched 2:3 and 3:4.

Due to the nature of the stimuli, ratio effects could only be analysed for fixed size and surface-area matched separately at Time 1 because different ratios were administered. Repeated-measures ANOVAs revealed a moderate effect in the surface-area matched condition, *F*(1,99) = 5.500, *p* = .021, η_p^2 = .05, and no effect in the fixed size condition, *F*(1,99) = .374, *p* = .54, suggesting that the performance on the fixed size ratios 3:4 (*M* = 10.76, *SD* = 2.72) and 5:6 (*M* = 10.58, *SD* = 2.29) was equal whereas children performed better on surface-area matched 2:3 (*M* = 11.42, *SD* = 2.42) ratio compared to 3:4 (*M* = 10.81, *SD* = 2.03). Only the ratio 3:4 was assessed in fixed size and surface-area matched, thus comparison between fixed size and surface-area matched could only be conducted for 3:4 ratio showing no significant difference in children's performance on ratio 3:4 depending on the size of squares (t(1,99) = .176, p = .86).

After Time 1, the ratios chosen to be investigated were 2:3, 3:4 and 5:6, and all three ratios were administered as surface-area matched and fixed size conditions. This permits an analysis of a 3 (ratio) x 2 (size) repeated-measures ANOVAs for each subsequent time point.

Analysis of the ratios of Time 2 showed main effects for ratio (F(2,232) = 63.46, $p < .001$, $\eta_p^2 = .354$) and size (F(1,116) = 25.401, $p < .001$, $\eta_p^2 = .108$), but no significant interaction. Children showed greater performance on fixed size trials (M = 12.13, SD = .16) than surface-area matched (M = 11.46, SD = .15), and significantly performing better on 2:3 ratio (M = 12.77, SD = .18), followed by 3:4 ratio (M = 11.77, SD = .18) and 5:6 ratio (M = 10.84, SD = .18).

Similar to Time 2, only main effects for ratio and size were significant at Time 3 (F(2,230) = 61.31, $p < .001$, $\eta_p^2 = .348$ and F(1,115) = 15.93, $p < .001$, $\eta_p^2 = .122$ respectively). Inspection of means suggests that fixed size stimuli (M = 11.33, SD = .42) were easier to solve than surface-area matched stimuli (M = 10.47, SD = .37). The results for ratio showed a pattern similar to Time 2, with significant differences between 2:3 (M = 12.44, SD = .45), 3:4 (M = 11.01, SD = .43) and 5:6 (M = 9.19, SD = .37).

Likewise, analysis of Time 4 showed main effects for ratio (F(2,218) = 126.605, $p < .001$, $\eta_p^2 = .537$) and size (F(1,109) = 41.077, $p < .001$, $\eta_p^2 = .274$). Children showed greater performance on fixed size trials (M = 13.56, SD = .38) than surface-area matched (M = 15.14, SD = .43), and the ratio 2:3 being the easiest (M = 16.79, SD = .49) compared to 3:4 (M = 14.46, SD = .44) and 5: 6 (M = 11.79, SD = .32; all ratio comparisons were significant).

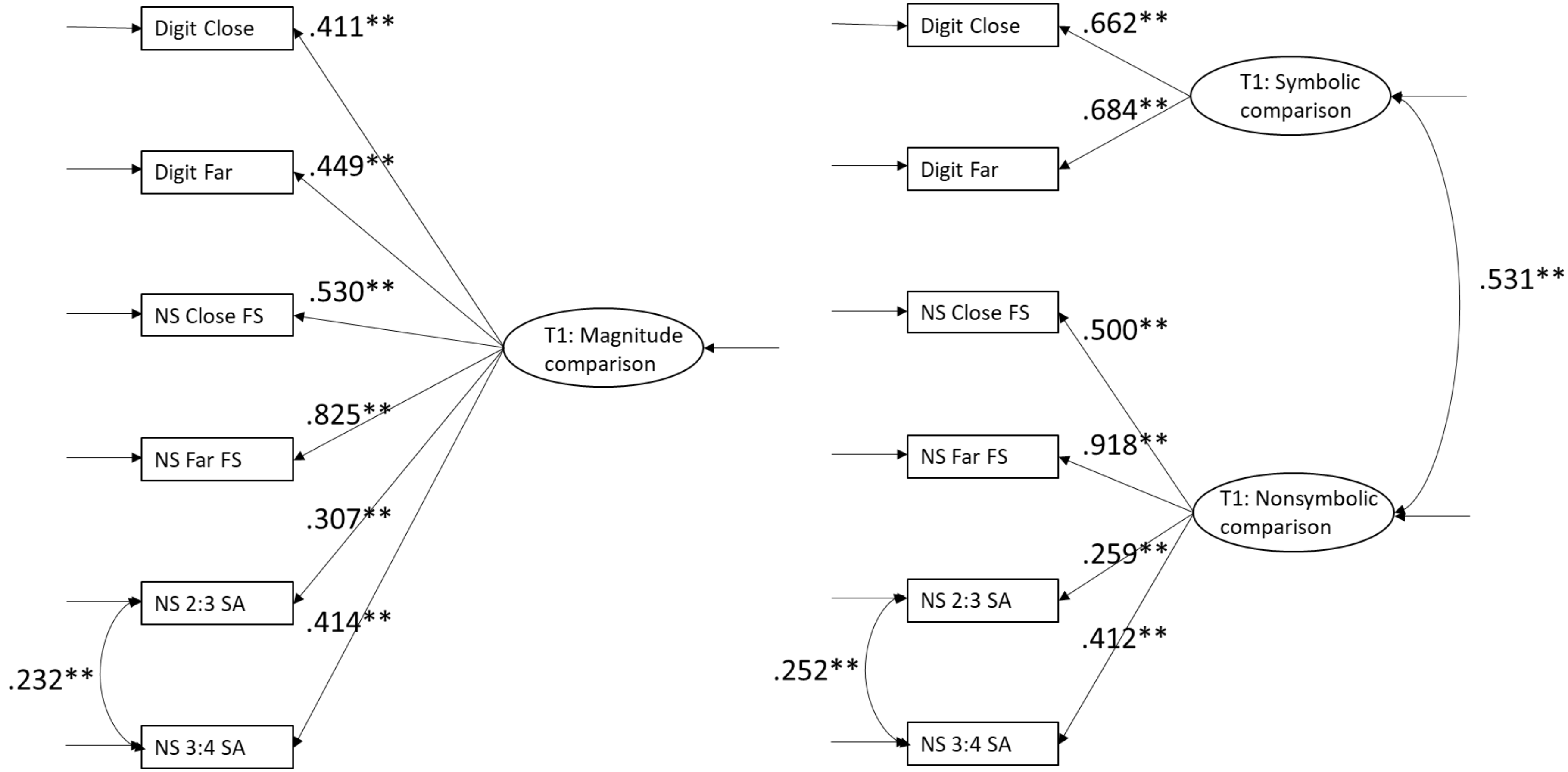
Analysis of Time 5 data revealed the same findings as in previous time points, with significant main effects for ratio (F(2,232) = 112.424, $p < .001$, $\eta_p^2 = .492$) and size (F(1,116) = 62.538, $p < .001$, $\eta_p^2 = .35$), but no significant interaction. Performance on fixed size (M = 17.09, SD = .49) stimuli was greater than surface-area matched (M = 15.29, SD = .45). Inspection of children's performance on ratio trials showed greater performance on 2:3 (M = 18.65, SD = .58) ratios compared to 3:4 (M = 16.63, SD = .51) and compared to 5:6 (M = 13.30, SD = .40).

3.2.4 Confirmatory factor analyses (CFAs) on comparison measures.

A series of confirmatory factor analyses were conducted to investigate the relation between the various magnitude comparison tasks. The CFAs examined whether magnitude comparison tasks are related by comparing a one-factor (general comparison ability) and a two-factor model (symbolic and nonsymbolic comparison). Furthermore, it was investigated whether this relationship changes over time and if the structure switches from a two-factor towards a unitary factor model or vice versa.

Based on the finding that the performance on Time 1 ratios fixed size 3:4 and 5:6 was equal whereas children performed better on surface-area matched 2:3 ratio compared to 3:4, the ratios for the fixed size manipulation were removed from subsequent analyses. To further simplify the model, the distance effect (close versus far) trials of the surface area matched manipulation were excluded. The chosen magnitude comparison tasks for analyses are: digit close, digit far, fixed size close, fixed size far and surface area matched ratio tasks.

3.2.4.1 Time 1. The first set of CFAs examined the nature of magnitude comparison tasks at Time 1, allowing the correlated error between surface-area matched ratios 2:3 and 3:4 because both tap into the ‘factor’ surface-area matched. Although all tasks loaded significantly on the single factor magnitude comparison in the one-factor CFA (Figure 3.2), the model did not provide an acceptable fit to the data, χ^2 (8) = 22.836, p =.004, $RMSEA$ = .136 (90% CI = .072 - .203), CFI = .827, $SRMR$ = .071, suggesting that a single factor is not sufficient and a better model would involve at least two factors (symbolic and nonsymbolic).



3.2. One factor (left side) and two factor (right side) CFA of magnitude comparison tasks (Time 1).

Figure 3.2 shows that the two-factor model provided an acceptable fit to the data, χ^2 (7) = 9.325, *p* = .23, *RMSEA* = .058 (90% CI = .000 - .144), *CFI* = .973, *SRMR* = .047. A chi-squared difference test confirmed that this model fited the data significantly better than the unitary model (χ^2_{diff} (1) = 13.511, *p* < .001). Inspection of the individual loadings revealed poor loadings for the difficult surface-area matched close and surface-area matched ratio 5:6 conditions, suggesting that these tasks are not sensitive enough and may be too difficult.

3.2.4.2 Time 2. A set of CFAs was conducted to assess the relationship between the magnitude comparison tasks at Time 2. A set of CFAs (Figure 3.3) was conducted using only the corresponding magnitude comparison subtasks from Time 1 (digit close, digit far, fixed size close, fixed size far and surface area matched ratios 2:3 and 3:4). The CFA included the correlated error between surface-area matched ratios 2:3 and 3:4 because both tap into the ‘factor’ surface-area matched.

The first model presenting a single factor did not provide an adequate fit to the data, χ^2 (8) = 25.295, *p* = .001, *RMSEA* = .136 (90% CI = .078 - .197), *CFI* = .881, *SRMR* = .067. However, the two-factor model provided a good fit to the data χ^2 (7) = 4.713, *p* = .695, *RMSEA* = .000 (90% CI = .000 - .087), *CFI* = 1.00, *SRMR* = .024 which was significantly better than the single-factor model (χ^2_{diff} (1) = 20.582, *p* < .001).

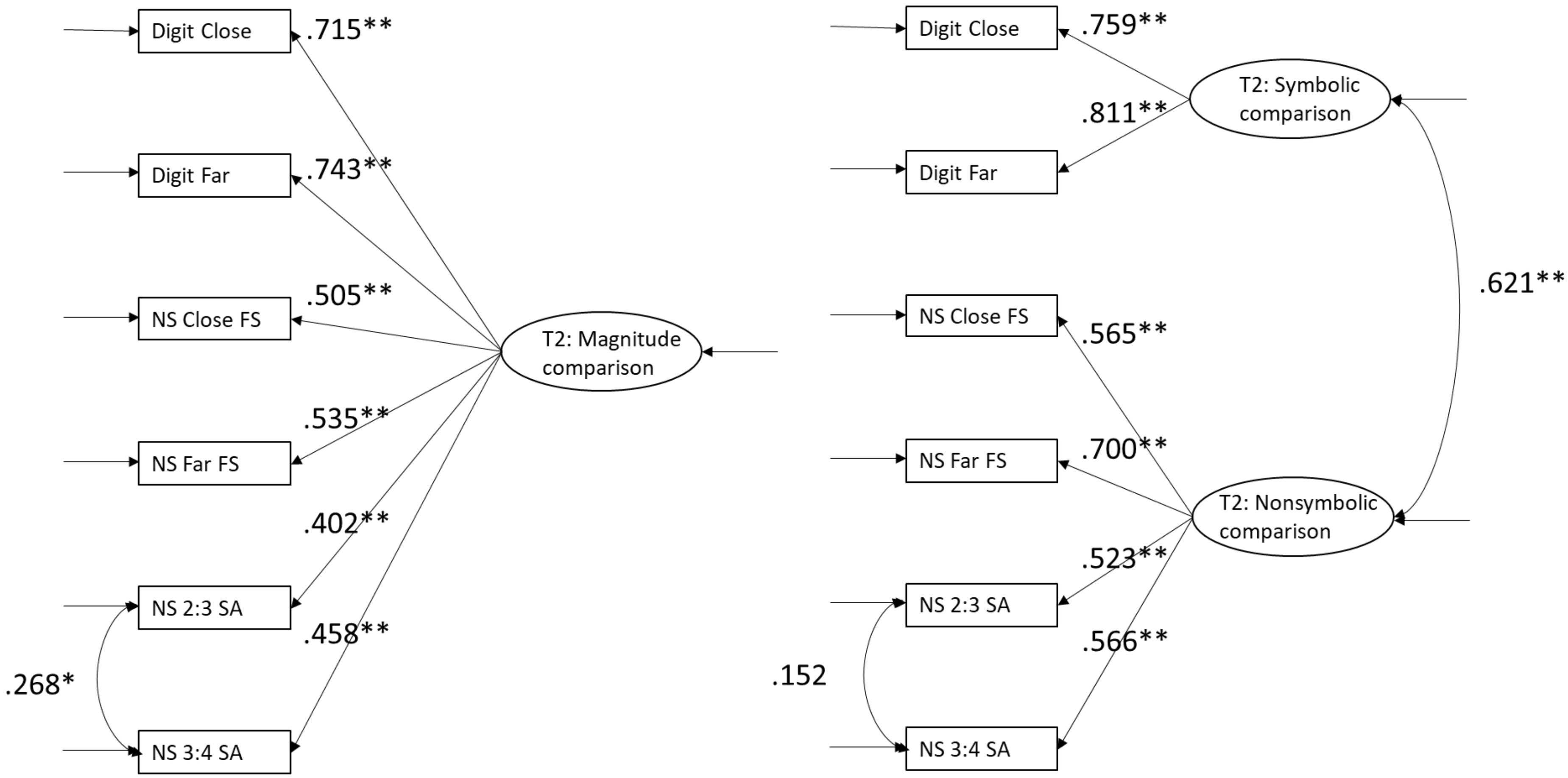


Figure 3.3. One factor (left side) and two factor (right side) CFA of magnitude comparison tasks (Time 2).

3.2.4.3 Time 3. The same set of CFAs as used in Time 2 were conducted to examine the relationship between the magnitude comparison tasks at Time. The models included the correlated error between surface-area matched ratio 3:4 and fixed size close because both tap into the ‘factor’ difficult-comparisons.

The first path model (Figure 3.4) investigates the construct of a single factor. This model provided an acceptable fit to the data (χ^2 (8) = 14.859, *p* = .062, *RMSEA* = .086 (90% CI = .000 - .153), *CFI* = .982, *SRMR* = .035). The two-factor model provided an even better fit to the data (χ^2 (7) = 6.705, *p* = .460, *RMSEA* = .000 (90% CI = .000 - .111), *CFI* = 1.00, *SRMR* = .022) compared to the single-factor model (χ^2_{diff} (1) = 8.154, *p* = .004).

At Time 3, the magnitude comparison tasks were administered as a time-constrained group test. Surface-area matched and fixed size conditions used the same subtasks, close versus far and ratios 2:3, 3:4 and 5:6. It was decided to investigate to what extent including all tasks may change the structure of the CFA. The first path model (Figure 3.5) investigates the construct of a single factor. This model provided an inadequate fit to the data, χ^2 (51) = 119.707, *p* < .001, *RMSEA* = .108 (90% CI = .083 - .133), *CFI* = .940, *SRMR* = .043. Not surprisingly, the two-factor model provided a better fit to the data, χ^2 (55) = 101.851, *p* = .197, *RMSEA* = .095 (90% CI = .068 - .121), *CFI* = .954, *SRMR* = .040; χ^2_{diff} (1) = 17.856, *p* < .001. Neither model using all magnitude comparison subtasks provided an optimal fit to the data despite the fact that multiple correlated errors were allowed to improve model fitness. Thus, further CFAs were run using only a chosen subset of the magnitude comparison tasks (based on Time 1; digit close, digit far, fixed size close, fixed size far and surface area matched ratios 2:3 and 3:4).

At this point, including all magnitude comparison subtasks worsened the model fit. Because of the small sample size, running complex CFAs with many manifest variables may reduce power of the analysis producing misleading results. In order to obtain acceptable model fitness, various correlated errors need to be included in the model (Figure 3.5). To prove this, a set of CFAs was run including all subtasks.

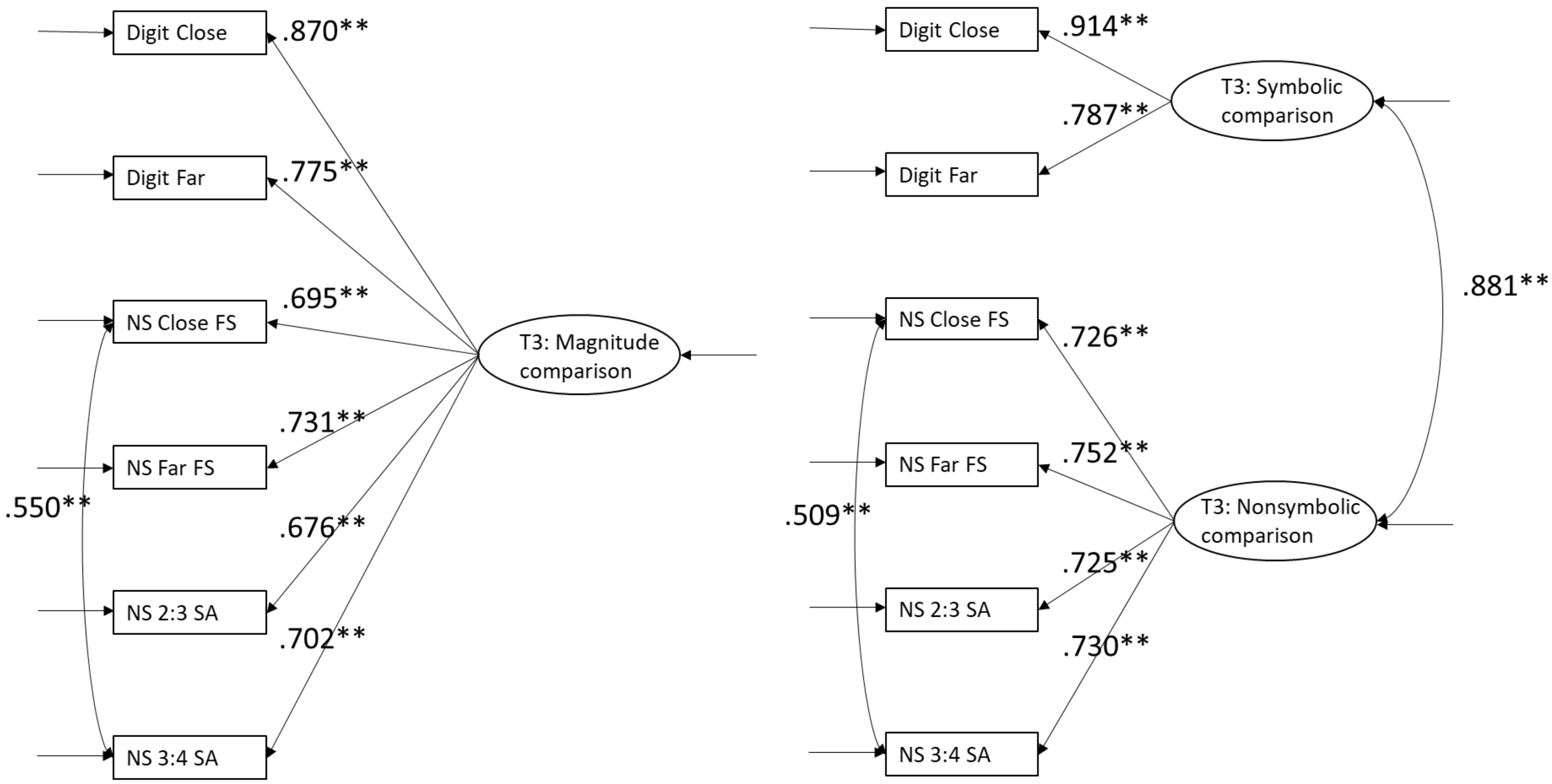


Figure 3.4. One factor (left side) and two factor (right side) CFA of magnitude comparison tasks (Time 3).

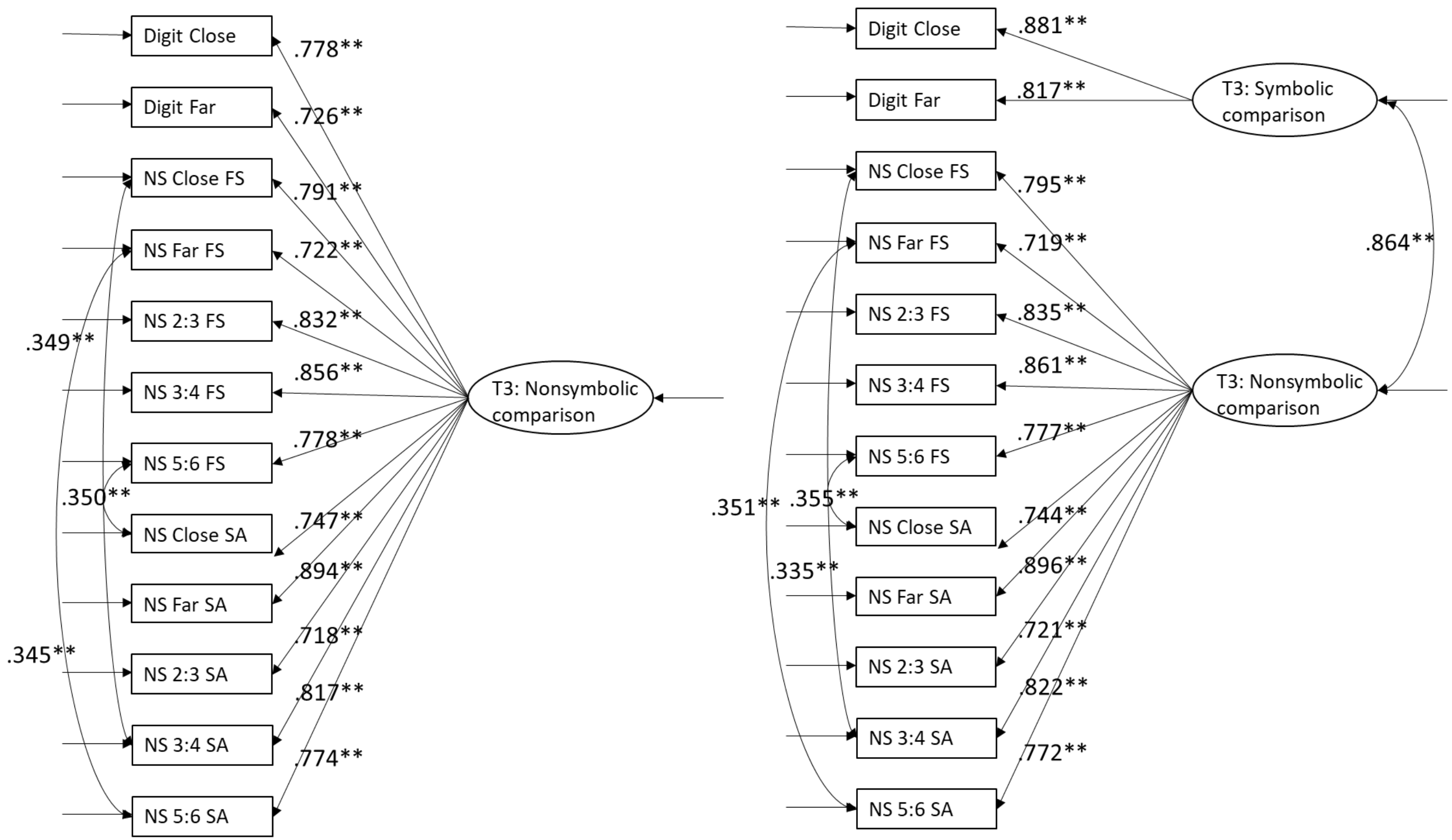


Figure 3.54. One factor (left side) and two factor (right side) CFA of magnitude comparison tasks using all subtasks (Time 3).

3.2.4.4 Time 4. Further sets of CFAs were run using tasks from Time 4 to assess the relationship of magnitude comparisons in the spring term of Year One. The models included the correlated error between surface-area matched ratios 2:3 and 3:4 because both tap into the surface-area-matched 'factor'.

A set of CFAs was conducted (Figure 3.6) using only the corresponding magnitude comparison subtasks from Time 1 (digit close, digit far, fixed size close, fixed size far and surface area a matched ratios 2:3 and 3:4). Both models provided a good fit to the data (single-factor: χ^2 (8) = 12.170, p = .144, *RMSEA* = .069 (90% CI = .000 - .141), *CFI* = .987, *SRMR* = .028 compared to two-factor model: χ^2 (7) = 11.639, p = .113, *RMSEA* = .077 (90% CI = .000 - .153), *CFI* = 986, *SRMR* = .028), but the difference between the two models was not significant (χ^2_{diff} (1) = .531, p = .466), thus the addition of a second factor did not improve the fit and the single-factor model is to be favoured.

3.2.4.5 Time 5. The last analyses on the development of magnitude comparison tasks comprises of a sets of CFAs (subtasks based on Time 1 model; Figures 3.7). The one-factor model provided an excellent fit to the data, χ^2 (9) = 13.320, p = .149, *RMSEA* = .064 (90% CI = .000 - .132), *CFI* = .991, *SRMR* = .023) and so did the two-factor model, χ^2 (8) = 10.916, p = .207, *RMSEA* = .056 (90% CI = .000 - .130), *CFI* = .994, *SRMR* = .021. This difference is not significant (χ^2_{diff} (1) = 2.404, p = .121) suggesting that adding a second factor did not increase the fit.

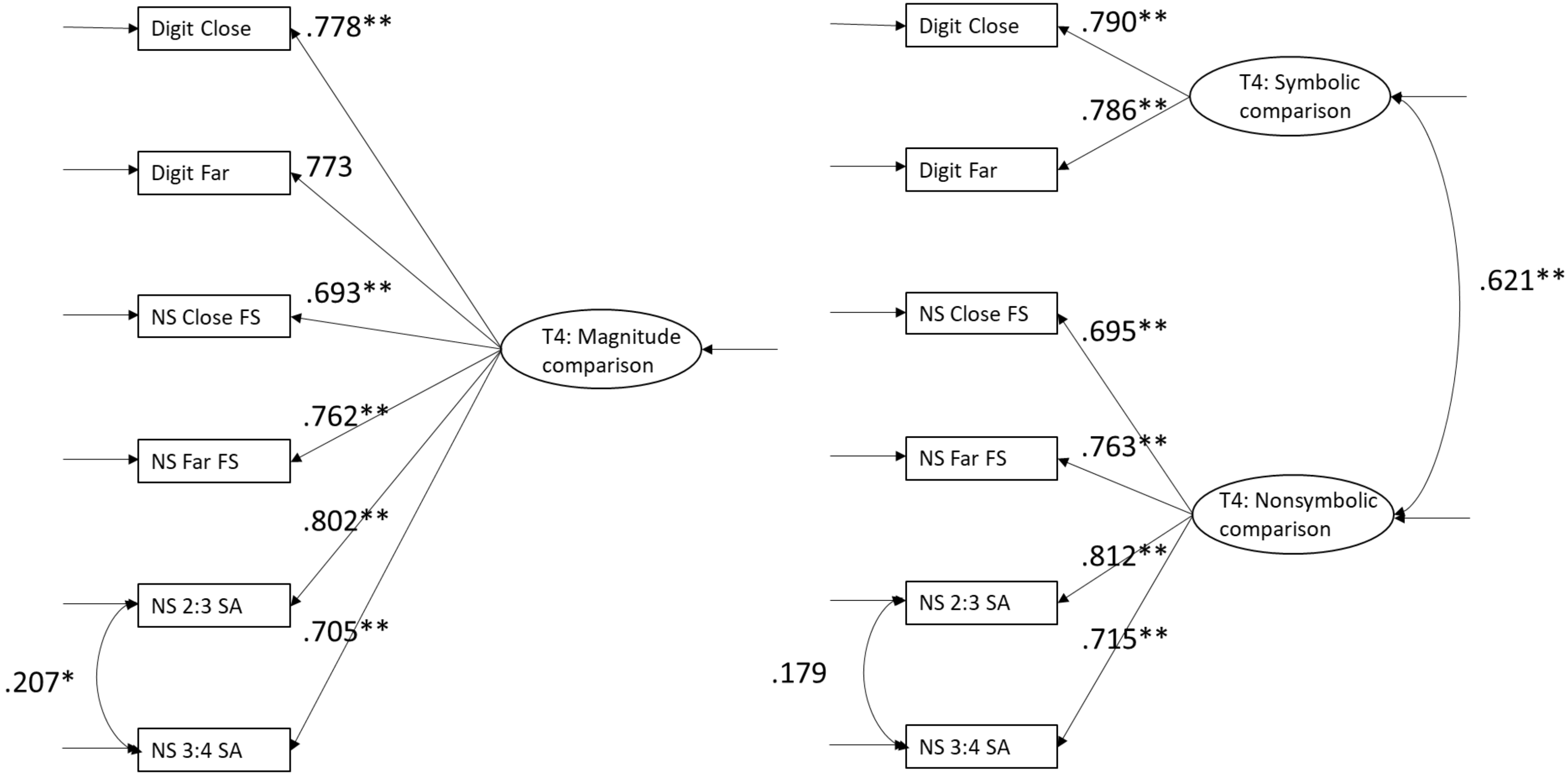


Figure 3.6. One factor (left side) and two factor (right side) CFA of magnitude comparison tasks (Time 4).

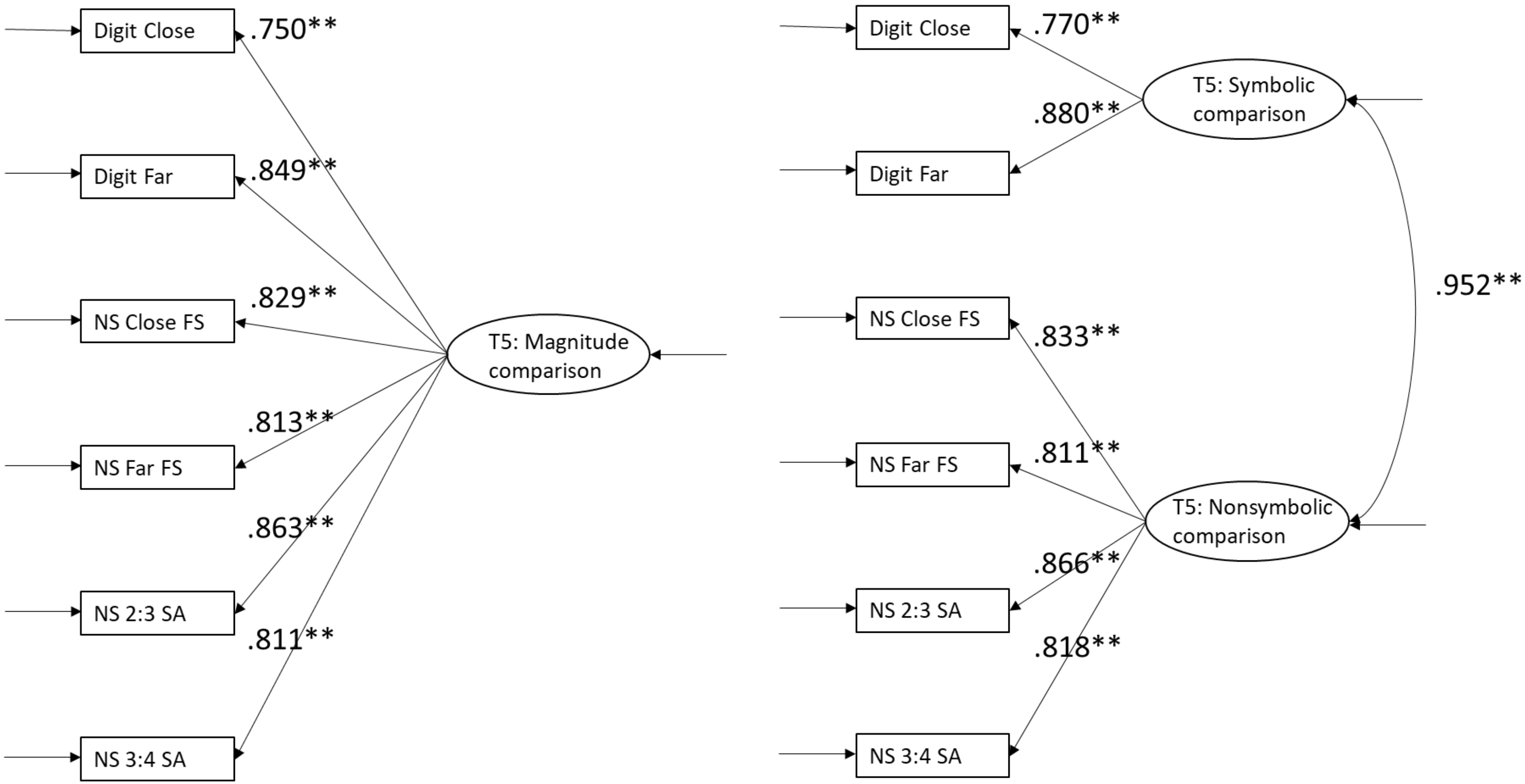


Figure 3.7. One factor (left side) and two factor (right side) CFA of magnitude comparison tasks (Time 5).

3.2.5.4 CFAs on comparison measures: investigating high achievers in number recognition.

As discussed earlier, while almost half of sample (45%; 45 children) scored the maximum on a single digit Arabic number reading task, the greater proportion of the sample made one or more errors. This may explain the Times 1 to 2 findings that symbolic and nonsymbolic comparison tasks are better explained by a two-factor than a one-factor model. Variability in symbolic comparison may register symbol knowledge rather than magnitude comparison, therefore the single magnitude comparison model fails to represent the data. Thus re-examining the data only using the scores of the high achievers on the number reading task (number wizards; children that scored the maximum possible score) could provide useful insight in the development and structure of magnitude comparison.

First, the CFAs for Time 1 were re-run using the same variables. The chosen magnitude comparison tasks for analyses are: digit close, digit far, fixed size close, fixed size far and surface area matched ratio tasks. Because neither of the surface area matched ratio tasks loaded significantly onto the single factor or the two-factor model, they were excluded from the analyses. The one-factor CFA model provided a moderate fit to the data, χ^2 (3) = 3.167, p = .367, *RMSEA* = .035 (90% CI = .000 - .256), *CFI* = .991, *SRMR* = .087. The two-factor model also provided an acceptable fit to the data, χ^2 (2) = 0.261, p = .878, *RMSEA* = .000 (90% CI = .000 - .146), *CFI* = 1.00, *SRMR* = .059. A chi-squared difference test confirmed that the two-factor model did not fit the data significantly better than the unitary model (χ^2_{diff} (1) = 2.906, p = .088).

Next, the CFAs for Time 2 were re-run. The variable fixed size close was removed because it did not load significantly onto the factors. The one-factor model fit was weak, χ^2 (5) = 13.713, p = .018, *RMSEA* = .230 (90% CI = .088 - .379), *CFI* = .838, *SRMR* = .085, however, the two-factor model fit was moderate, χ^2 (4) = 9.017, p = .061, *RMSEA* = .195 (90% CI = .000 - .367), *CFI* = .906, *SRMR* = .069. The difference between the two models was significant, diff: χ^2_{diff} (1) = 4.696, p = .030.

Likewise, the Time 3 results revealed that the single-factor model, χ^2 (9) = 20.345, p = .016, *RMSEA* = .193 (90% CI = .079 - .305), *CFI* = .924, *SRMR* = .048, as well as the two-factor model, χ^2 (8) = 15.350, p = .053, *RMSEA* = .164 (90% CI = .000 - .288), *CFI* = .950, *SRMR* = .062, provided an adequate fit to the data. Again, the two-factor model fit was significantly different from the one-factor model, χ^2_{diff} (1) = 4.995, p = .025.

Re-running the CFAs for Times 4 and Time 5 confirmed the previous findings using the whole sample. At Time 4, the two-factor model providing an excellent fit to the data, χ^2 (8) = 5.667, p = .6.85, *RMSEA* = .000 (90% CI = .000 - .162), *CFI* = 1.00, *SRMR* = .024, was not significantly different from the single-factor model, χ^2 (9) = 5.667, p = .773, *RMSEA* = .000 (90% CI = .000 - .136), *CFI* = 1.00, *SRMR* = .024, χ^2_{diff} (1) = 0, p = 1.00. At Time 5, the unitary model provided an excellent fit to the data, χ^2 (9) = 12.639, p = .180, *RMSEA* = .111 (90% CI = .000 - .241), *CFI* = .979, *SRMR* = .030, as did the two-factor model, χ^2 (8) = 11.610, p = .170, *RMSEA* = .117 (90% CI = .000 - .253), *CFI* = .979, *SRMR* = .029. The difference was not significant, χ^2_{diff} (1) = 1.029, p = .310.

3.3 Conclusion.

This chapter focused primarily on the nature and development of the ANS measured using magnitude comparison tasks. For this purpose, distance and ratio effects were first investigated followed by a detailed analysis of the structure of magnitude comparison tasks and its change over time using confirmatory factor analyses.

Overall, the comprehensive analyses of children's performance on symbolic and nonsymbolic magnitude comparison tasks revealed three findings: First, children's performance on magnitude comparison tasks generally showed significant distance and ratio effects for both symbolic and nonsymbolic comparisons with better performance on the far trials than close confirming previous findings (Barth et al., 2003; Piazza et al., 2010; Xu and Spelke, 2000; Halberda et al., 2008; Gilmore et al. 2010; Libertus et al., 2011; Mazzocco et al., 2011; Halberda and Feigenson, 2008). There was a significant interaction for the symbolic distance effect across Times 1 and 2. Children performed significantly better on far trials than close trials at Time 2, but the distance effect for the symbolic task at Time 1 (numeric distance effect) was not significant. This may be due to the fact that some children had difficulties reading the Arabic numerals. It was noted that a third of the children made at least two mistakes in reading single digit Arabic numerals. If mastery of the single digit Arabic numerals is taken into account, a marginal distance effect can be found even in young children. As expected, no such limitation applied to performance on nonsymbolic comparison.

However, the findings for nonsymbolic comparison included two interactions: at Time 1 where the manipulation of square size affected far trials (easier items in general) and at Time 5 when square size influenced the performance of items in the close condition (harder condition). The findings suggest that manipulating square size affects very young children differently than older children suggesting that exposure to the task as well as a developmental improvement in magnitude comparison skills boost children's performance. These interactions may support the findings from Sekuler and Mierkiewicz (1977) that fourth and seventh graders slope of the function relating to judgement time to distance was comparable to adult performance, whereas the function of kindergarten and first grade children was much steeper. The authors concluded that there are no qualitative differences supported by the fact that the shape of the numerical difference effect was the same for the groups, but rather quantitative differences. According to the authors, a steeper slope may imply either that the representation of numerical magnitudes is compressed in younger children, or that discriminal dispersion around the means are larger in young children, or a combination of the two.

However, Sekuler and Mierkiewicz (1977) used only symbolic comparison tasks. The nonsymbolic interaction may be more complicated. These interaction may be due to the perceptual advantage of fixed size stimuli over surface-area matched stimuli, or the requirement to supress incongruent stimuli in fixed size condition may affect children differently at different ages according to the difficulty of the comparison (far versus close).

Second, the results revealed nonsymbolic ratio effects (2:3 > 3:4 > 5:6) across all time points. Children performed more accurately on ratios with a large difference (i.e. 2:3) than ratios with a small difference (5:6). Furthermore, manipulating the feature size impacts children's performance on comparison tasks. Fixed size arrays were generally easier to discriminate for both, distance and ratio trials, than surface-area matched arrays suggesting that it is more difficult for children to ignore the prominent feature size in the surface-area matched condition where the array with fewer stimuli has bigger squares compared to many tiny squares. Previous studies have shown that performance on ANS measures increases with age, with adults discriminating numerosities outside of the ability of infants. Halberda and Feigenson (2008) identified the Weber fraction of ANS in three-, four-, five- and six-year-old children and adults using dot arrays ranging from 1 to 14 dots spanning the ratios from 1:2 through 9:10. They further controlled for object area similar to the current fixed size versus surface-area matched manipulation. The results showed an increase in performance over time with six year-olds performing like adults. The current study found a comparable increase in performance over time.

Third, the dynamic relation between symbolic and nonsymbolic magnitude comparison tasks changes over time. This change coincides with children's entry to the formal school system. Symbolic and nonsymbolic comparison tasks loaded on separate factors at four to five years of age. Interestingly, the same pattern emerged at Time 3, with magnitude comparison tasks being represented by two separate underlying factors (symbolic and nonsymbolic) rather than one general comparison factor. However, the distinction between the factors is declining and it seems that children's representation and processing of magnitude comparison tasks at the age of 5 years and 6 months (autumn term of Year One) is changing towards the general comparison ability construct. To further investigate this hypothesis, analyses of the subsequent two time points are crucial. If this hypothesis is true, a shift towards the single-factor model should occur. At Times 4 and 5, the single-factor model should be preferred meaning that magnitude comparison tasks load on one general comparison factor and not two distinct factors (symbolic and nonsymbolic) confirming the developmental trend towards a general magnitude comparison factors.

Children's pre-school representation of magnitude comparison tasks may best be described by two distinct underlying factors: symbolic and nonsymbolic magnitude comparison (Libertus et al., 2011, Piazza, 2010; Piazza and Dehaene, 2004).

This distinction is vanishing slowly around Year One moving from two constructs towards one general comparison ability construct (see also Kolkman et al., 2012). It seems that the shift in the processing of magnitude comparison tasks may be complete by the end of Year One (6 years and 4 months of age). Questions remain on why this change in the representation and processing of the magnitude comparison occurs.

One possibility would be that children's mastery of the Arabic numeral system and their understanding of magnitude in general may play a crucial role in the development of comparison tasks. Interestingly, the change appears around the time after children entered school and are formally trained in numeracy. Also, after five testing sessions, children were very familiar with the task and stimuli and this exposure may further foster the change in processing. The findings suggest that at an early, pre-school age, processing of magnitude comparisons may load heavily on cognitive resources, and that children devise different strategies to solve symbolic and nonsymbolic magnitude tasks. Through exposure and formal training on numeracy the processes become more automatized and rely on broader general comparison abilities.

This hypothesis is supported by findings on high achievers on number reading task. These number wizards achieved the maximum score on the number reading task at Time 1 (four years of age). If using only data from the number wizards, then the two-factor and single-factor models do not significantly differ, suggesting, according to the principle of parsimony, that the latter is the better model and that the performance of number wizards can best be explained by one general magnitude comparison construct at Time 1. At Times 2 and 3 however, the two models do slightly differ favouring the two-factor model, though the significance alpha level was only .05. Nevertheless, these findings point to the fact that children's understanding of numerals may be mediating the divide between symbolic and nonsymbolic magnitude comparison. The Times 3, 4 and 5 results are similar to the findings from the whole sample that symbolic and nonsymbolic magnitude comparison form one general magnitude comparison construct. All in all, the findings indicate that children's mastery of Arabic numerals plays a crucial part in the structure of the ANS. Once children have a complete understanding of the single digit numerals, then the symbolic and nonsymbolic comparison task are best described by a one-factor model. These findings will inform future studies which should also account for children's number recognition skills when examining the relationship between ANS and arithmetic.

In summary, the results clarify the little investigated structure of symbolic and nonsymbolic magnitude comparison. At pre-school age, ANS tasks show two distinguishable skills compared to the integration of the ANS skills into one general magnitude comparison structure. In view of these recent results, previous findings on the relationship between symbolic and nonsymbolic comparisons and their impact on arithmetic skills at school age should carefully be re-examined. The following chapters will investigate the concurrent as well as longitudinal prediction of arithmetic focusing on the role of magnitude comparison.

Chapter 4. Concurrent Prediction of Early Arithmetic across Time.

Studies indicate that language deficits (SLI) may affect a wide range of numeracy skills differently (Donlan et al., 1998; Donlan, and Gourlay, 1999; Fazio, 1994; 1996). Children with SLI performed lower in rote counting than typically developing children of the same age (Donlan et al., 2007). Furthermore, Cowan et al. (2005) and Donlan et al. (2007) found that difficulties in producing the spoken number sequence, as well as poor comprehension of language, are significantly associated with calculation. Kleemans et al. (2011, 2012) found a relationship between grammatical ability and early numeracy skills. However, the relationship between these skills is complex, and runs counter to other findings which indicate independence between verbal and nonverbal calculation skills (Nunes and Bryant, 1996; Jordan et al., 1994).

The role of ANS is still debated. Evidence in support of the importance of the ANS comes from correlational studies showing that individual differences in ANS and general mathematical achievement are strongly correlated (Halberda et al., 2008; Gilmore et al. 2010; Libertus et al., 2011; Mazzocco et al., 2011; Halberda et al., 2008. In contrast, some studies have failed to report a significant relation between nonsymbolic ANS measures and arithmetic (Holloway and Ansari, 2009; Iuculano et al., 2008; Sasanguie et al., 2012; Kolkman et al., 2012; Vanbinst et al., 2012).

Recent longitudinal studies produced mixed results (Desoete et al., 2012; Lyons et al., 2014). A recent study by Göbel et al. (2014) addressed the relation between nonsymbolic and symbolic judgement tasks and their role as longitudinal predictors of arithmetic development in six-year-olds. The authors reported that symbolic and nonsymbolic magnitude comparisons define a unitary factor, which was a strong longitudinal correlate of arithmetic skills. The path model revealed that number identification task, in which spoken numerals were presented to be matched to the corresponding Arabic numeral, was the most powerful longitudinal predictor of arithmetic skills at age seven, apart from the auto-correlate.

This chapter aims to identify the concurrent predictors of perfromance in arithmetic during a two year period at the sensitive transition from pre-school to formal school, when there is a rapid change in the development in basic arithmetic skills.

4.1 Methods.

4.1.1 Participants.

The same participants were used as described in Chapter 2 (p. 42)

4.1.2 Materials.

Children were assessed on the following measures.

4.1.2.1 Measures taken at Time 1.

Nonverbal intelligence. Nonverbal intelligence was assessed using a traditional matrix reasoning task. Set A of the Raven's Coloured Progressive Matrices (Raven's CPM; Raven et al., (1993)) was chosen. Items were administered according to the manual. Children were given an incomplete matrix puzzle and asked to choose from six missing pieces to mark the piece that completes the matrix. Three novel practice items were administered before the test trials. These were created based on the features of the original matrices. One point was given for each correct response with a maximum possible score of 12.

General Language Knowledge.

Grammatical ability. The children's grammatical ability was assessed using the Test for Reception of Grammar II (TROG-2; Bishop, 2003). The TROG-2 was in booklet form and consisted of twenty blocks with four different items. Most TROG-2 items had one target picture, one syntactic distractor and two lexical distractors. Complete sentences were read aloud, using the correct stress pattern for each experimenter, for each item, for each child. The children had to point to one of four pictures. A child failed the block if one item was incorrect. Testing was terminated when the child failed five consecutive blocks. The raw scores (number of blocks passed) were reported.

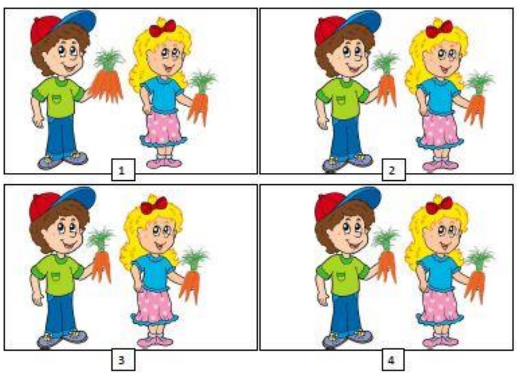
Vocabulary. Children's vocabulary skills were examined using the British Picture Vocabulary Scale 3rd Edition (BPVS - III; Dunn et al., 2010) consisting of thirteen sets each with twelve items. The BPVS – III was executed according to the manual. Starting and termination points were identified following the manual. Children identified which of four pictures best matched the spoken target word. The

BPVS – III is highly reliable (reported Cronbach’s α = .91). The raw scores (number of correct responses) were reported.

Specific math-related language ability. Furthermore, testing involved a new task assessing children’s understanding of quantitative relations. The Test of Relational Comprehension (TRC; Figure 4.1; see Appendix 2 for list of items), designed by Chris Donlan, addressed the issue of mathematical language, testing understanding of relational terms such as *more* or *less*.

The test was originally designed for older children and included the more complex concept of *less*. Investigating nursery children, the test was re-designed for younger participants, assessing only the relational concept of *more*. Specific testing addressed children’s comprehension of quantitative statements over mass nouns, X has more (noun) than Y, and comparative adjectives with *more* (X is more beautiful, more handsome, more colourful or more comfortable than Y). The TRC uses a four-picture selection format (similar to TROG above) with three distractors.

The test was divided into two parts; an easy and a hard part. Both parts had each three items for countable nouns, mass nouns and comparative adjectives. The difference was that the former consisted of the target and three identical distractors whereas the latter consisted of the target and three variations of distractors (reverse pattern, both objects having the same amount – same amount of the lower and upper end of the targeted quantity; see Figure 2). Complete sentences were read aloud and the children had to point to one of four pictures. A maximum of 24 points could be achieved, 12 for each part. The number of correct items was counted. The child’s response was noted.

a) 

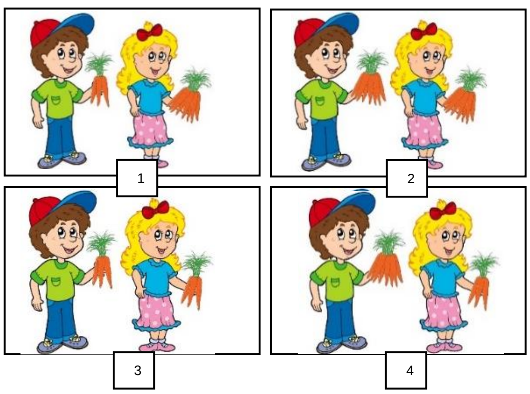
b) 

Figure 4.1. Example TRC item “The boy has more carrots than the girl” in the a) easy condition and b) the hard condition.

Transcoding. To estimate children's number knowledge, a variety of tests, based on understanding of the Arabic numeral system, was conducted.

Number Identification. Children were presented with four Arabic numerals and were asked to point to the numeral that matched the spoken number among three distractors which were chosen to reflect common errors that young children make (Figure 4.3; Appendix 4 for list of items). Based on Mix et al. (2014), targets not only included units but also tens and hundreds (target numbers were 6, 28, 206, 7, 91, 2, 41, 52, 11, 69, 37, 43, 74, 168, 13 and 85).

206	260	26	2060

Figure 4.2. Example Number Identification task "Can you point to number 206."

Number Writing. Similar to the letter writing task, we asked the children to transcribe twelve Arabic numerals (2, 9, 7, 4, 8, 10, 6, 1, 20, 3, 100 and 5) which were presented verbally. Two points were awarded for each numeral (accuracy and orientation; based on the *Letter Writing* task of Caravolas, Lervåg, Mousikou, Efrim, Litavský, Onochie-Quintanilla, Salas, Schöffelová, Defior, Mikulajová, Seidlová-Málková and Hulme, 2012), thus resulting in a maximum score of 24.

Reading Arabic numerals. Knowledge of Arabic numerals was assessed using a Numeral Reading task in which children had to read out loud the Arabic numerals (MS Office 2013, Comic Sans MS, size 350) one to ten. The numbers were presented in random order to avoid any effects of number sequence knowledge.

Rote Counting. Children were asked to count from one forward. Testing stopped after the child reached the number 111, did not know how to count farther, or if they showed signs of distress. The highest number counted without making mistakes was reported.

Magnitude Comparison. Various symbolic and nonsymbolic comparison tasks were created for the study, based on those by Göbel et al. (2014). Each comparison pair was presented on a single page. Children were given one point for every correct comparison with a maximum score of 16 for each version and 160 overall (Figure 4.2; for more details see Chapter 3, p. 50).

Arithmetic Skills. The children's basic arithmetic skills were assessed using simple addition problems (Appendix 9). The test comprised ten simple additions with sums less than ten (1 + 3; 2 + 1; 2 + 2; 1 + 4; 3 + 1; 1 + 5; 2 + 3; 1 + 6; 3 + 3; 4 + 4). All arithmetic problems were presented in Arabic notation (MS Office 2013, Comic Sans MS, size 260) and, simultaneously, in spoken form most familiar to the child. Problems were arranged so that additions with same sums or similar summands were never adjacent. Children were encouraged to use wooden sticks provided or their fingers if needed. The preferred method of referring to additions ("add" or "plus") was determined by asking the teachers. Before the main testing, two practice problems (1 + 1, 1 + 2) were administered. Testing was terminated early if a child showed signs of confusion or lack of concentration. The maximum score was ten.

4.1.2.2 Measures taken at Time 2.

Specific math-related language ability. There was a ceiling effect with the easy part of TRC at Time 1, with 73% of the sample scoring nine out of ten or higher. Hence, the hard part of the TRC seemed to be the more sensitive measure of understanding of *more* at that age and therefore the easy part was dropped. To enhance the sensitivity of the hard part, items of the easy part were re-designed to match the hard part and half of the items were randomly chosen and re-configured as *less* trials. In the *less* trials children were asked to point to the picture that goes with e.g. 'the boy has less pasta than the girl'. Testing procedure was the same as Time 1 (see above for more details and Appendix 3). There were 12 *more* sentences and 12 *less* sentences, giving a maximum score of 24.

Transcoding. Although the difficulty level had to be adjusted for most number-knowledge tasks due to ceiling or close-to-ceiling effects, testing procedures did not change (see Time 1 for more details).

Number Identification. The following numbers were target numbers at Time 2: 6, 28, 206, 7, 91, 356, 2, 41, 52, 11, 69, 37, 807, 43, 74, 168, 13, 670, 614 and 85 (Appendix 5).

Number Writing. Children were asked to write down Arabic numerals (same targets as Time 1: 2, 9, 7, 4, 8, 10, 6, 1, 20, 3, 100 and 5) which were presented verbally. Contrary to Time 1, only one point was awarded for each numeral, thus resulting in a maximum score of 12.

Reading Arabic numerals. Based on children's performance at Time 1, ten more numbers were included in the number reading task. The following twenty numbers were administered in random order: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 19, 20, 100, 150, 210 and 437.

Rote Counting. The same task was used as in Time 1 (see above for more details).

Magnitude Comparison. The same symbolic digit and nonsymbolic magnitude comparison tasks were used and administered as in Time 1 (see above for details).

Arithmetic Skills. Due to the ceiling level performance at Time 1, the following adjustments were made to the basic calculation task (Appendix 10):

Two parallel forms of the task were created which comprised of ten simple additions with sums less than ten (both forms were equal in difficulty level; Form A: 1 + 3; 2 + 1; 1 + 5; 2 + 3; 4 + 5; 7 + 2; 3 + 5; 4 + 2; 5 + 2 and 2 + 6; Form B: 1 + 4; 3 + 1; 2 + 5; 4 + 2; 1 + 6; 3 + 6; 2 + 7; 6 + 2; 4 + 3 and 3 + 5). To raise the sensitivity of the task further, children were given only three minutes to solve as many problems as possible. The two forms were given in two separate testing sessions. The order of presentation of the forms was counterbalanced. The total number of correctly solved problems was recorded.

4.1.2.3 Measures taken at Time 3

Transcoding. All number knowledge tasks, except for number writing, were individually administered.

Number Identification. Children were asked to identify the following numbers among distractors: 6, 28, 206, 70, 91, 356, 50, 41, 52, 11, 69, 37, 807, 43, 74, 168, 13, 670, 614 and 85 (Appendix 6).

Number Writing Children were asked to write down the following Arabic numerals (12, 9, 73, 4, 18, 10, 16, 146, 21, 30, 100, 5, 500, 308 and 754). Testing and scoring procedure was the same as Time 2.

Reading Arabic numerals. Children were asked to read out loud the following twenty numbers: 8, 16, 201, 12, 20, 55, 9, 13, 100, 150, 14, 15, 11, 31, 210, 60, 437, 19, 10 and 142.

Rote Counting. It was assumed that most children have mastered the traditional rote counting procedure (as used at Time 1). Thus three new counting tasks were used were the children had to count from a given number until the experimenter asked them to stop. The three tasks were counting from one to 40, from 94 to 110 and counting backwards from 25. Similar to the Times 1 and 2 task, testing stopped after the child reached the target number, did not know how to count farther, or if they showed signs of distress. The furthest number counted in each sequence without making mistakes was reported for all three tasks.

Magnitude Comparison. A recent study by (Göbel et al., 2014) showed that children in Year One can successfully perform magnitude comparison tasks in a group setting. Thus the magnitude comparison task used in this study was redesigned as a group test using the same stimuli pairs created at Times 1 and 2. Symbolic and nonsymbolic comparisons were presented in pairs of two adjacent 2.1 cm x 2.1 cm boxes. Children were asked to tick the bigger number or box with more dots (for more details see Chapter 3, pp. 50).

Arithmetic Skills. *Fluency*. Children's speeded arithmetic skills (fluency) was assessed using the 'addition' and 'addition with carry' subtests of the Test of Basic Arithmetic and Numeracy Skills (TOBANS; Brigstocke et al., 2016). Children were asked to complete as many arithmetic problems as possible in one minute. In the 'addition' subtask, children were presented with simple addition problems with sums less than ten and in the 'addition with carry' subtask the sums were bigger than ten but smaller than twenty. One point was awarded even if the numeral was written backwards (maximum score $_{\text{addition}}$ = 90; maximum score $_{\text{addition with carry}}$ = 30). This task was administered as a group task (Appendices 11 and 12).

4.1.2.4 Measures taken at Time 4.

Working memory.

*Central Executive Functioning***.** To assess children's selective attention, each child completed a Visual Search task (Appendices 17 and 18). Children were asked to cross out as many red apples possible in one minute and ignore the distractors - red strawberries and white apples. To familiarise the children with the task, they were presented with pictures of all stimuli and were then asked to point to a particular one. Practice trials were administered beforehand. First, children were asked to complete the easy version A with relatively big, easy to distinguish stimuli. The number of correctly identified targets (17 red apples) and correctly rejected distractors (36 white apples and 37 red strawberries) were reported as well as the number of missed targets and wrongly marked distractors in order to calculate each child's d' (d prime; for exact calculation of d' see Appendix B). In addition, we further assessed a harder version because of the easy discrimination of items of form A which may be insensitive at the older age range. The target and distractor items of the harder version B were smaller and more difficult to discriminate (30 target red apples, 135 white apples and 135 red strawberries). Similar to the easy version, the children had one minute to find as many red apples as possible. Both versions were administered in the group session.

Transcoding. All number knowledge tasks, except for number writing, were individually administered.

Number Identification. Target numbers were: 6, 28, 206, 70, 91, 356, 50, 41, 52, 11, 69, 37, 3013, 807, 43, 74, 168, 13, 670, 614, 85, 819, 1109, 617 and 1220 (Appendix 7).

Number Writing. Children were asked to write down the following Arabic numerals: 12, 19, 73, 14, 18, 10, 16, 146, 21, 30, 100, 15 and 207. Testing and scoring procedure was the same as Time 2.

Reading Arabic numerals. Children were asked to read out loud the following numbers: 8, 16, 201, 12, 20, 55, 9, 13, 100, 150, 14, 15, 11, 31, 210, 60, 437, 19, 10, 142, 1109, 617, 1220, 819 and 2212.

Magnitude Comparison. The same tasks as at Time 3 were used (see above for details).

Arithmetic Skills. *Fluency.* In addition to the tasks at Time 3 (addition and addition with carry), children were also presented with the 'subtraction' subtask (Brigstocke et al., 2016). Similar to addition, children were asked to solve as many of the 90 subtraction problems as possible in one minute (Appendices 11-13).

4.1.2.5 Measures taken at Time 5.

Working memory.

*Executive Functioning***.** Similar to previous testing points, the visual search task was executed to assess children's selective attention. However, only the hard version with small stimuli from Time 4 was used. Additionally, a second version of this task was created with the same number of targets and distractor, just in a different, random order.

Literacy.

*Reading Skills***.** Children's reading skills were assessed using the Test of Word Reading Efficiency–Second Edition (TOWRE–2). The TOWRE-2 assesses children's ability to pronounce printed words (Sight Word Efficiency) and decode phonemically regular nonwords (Phonemic Decoding Efficiency) accurately and fluently, but only the nonwords subtask was used in the analysis. Children were asked to read as many nonwords as possible within 45 seconds. Eight test items were presented in vertical lists prior to the test list of 66 nonwords. The TOWRE-2 has four alternative forms (A through B). Only form A was administered individually. One point was awarded for each correctly decoded nonword.

Spelling Skills. Children's spelling skills were assessed using the Single Word Spelling Test (SWST). In this group test, children were asked to write down 30 high frequency words increasing in difficulty. First, the experimenter read out the whole sentence before repeating the target word only. All 30 items were administered making sure that all children were on the same item. One point was awarded for each correctly written word.

Transcoding. All Number knowledge tasks, except for number writing, were individually administered.

Number Identification. The following numbers were the difficulty-adjusted target numbers: 6, 28, 206, 70, 414, 91, 356, 50, 41, 7014, 52, 11, 69, 37, 528, 3013, 807, 4807, 43, 74, 168, 713, 13, 670, 614, 952, 85, 819, 1109, 617, 1220 and 493 (Appendix 8).

Number Writing Children were asked to write down the following Arabic numerals (12, 19, 73, 97, 14, 18, 113, 10, 16, 146, 4107, 21, 30, 100, 366, 15, 207, 1023 and 291). Testing and scoring procedure was the same as Time 2.

Reading Arabic numerals. Children were asked to read out loud the following numbers: 8, 16, 201, 12, 20, 309, 55, 9, 13, 100, 544, 150, 14, 15, 956, 11, 31, 210, 3614, 60, 437, 19, 10, 142, 387, 1109, 617, 1220, 819, 2212, 4097 and 438.

Magnitude Comparison. The same tasks as at Time 3 were used (see above for details).

Arithmetic Skills. *Fluency.* The same tasks as at Time 4 were used (see above for details).

4.1.3 Procedure.

The comprehensive test battery was divided into 20 to 40-minute-blocks at each time point to counterbalance effects or order such as learning and motivational effects. The testing order within each block was counterbalanced. Testing was carried out five times over a 25-month period from the summer term of nursery (May-June 2014) through to the summer term of Year One (June 2016). Wherever possible, each child was seen by the same experimenter. The main researcher was assisted by several research assistants from undergraduate psychology classes. They were trained on how to administer the test battery and were given instructions on how to work with young children. Children were tested individually at Times 1 and 2, and at Times 3, 4 and 5 individually or in small groups in a separate room or another quiet place in the school. The tasks that were tested individually after Time 3 included math-related language comprehension, number reading, number identification and counting. Each child met with the experimenter ideally two to four days in a row, depending on the number of blocks, to avoid lack of motivation or concentration. If testing in groups, the ratio of experimenters to children was 1:3.

Preliminary to the testing, the experimenters attended at least one day in each class so that the children got to know them and felt more comfortable around them. The experimenters told each child that they would play games and asked questions such as “How are you?” or “How old are you?”.

All unstandardized tests included practice items. Concerning feedback, children received only concrete feedback on their performance for practice items and general praise and encouragement throughout the tests.

4.2 Results.

The focus of the analysis was to identify concurrent predictors of early arithmetic and examine how the prediction from language, ANS and children’s understanding of the Arabic numeral system changes across time from pre-school to the conclusion of the first year of formal schooling. To answer this research question concerning the concurrent predictors of early arithmetic, descriptive statistics are presented first followed by a set of SEM path models estimated with Mplus Version

7 (Muthén and Muthén, 2013). The dependent variable was the latent arithmetic factor at each testing point and independent variables consisted of the latent factors measured at each testing point. Correlated residual errors of the manifest variables were included if this residual covariance improved the goodness of fit of the model significantly. Only theoretically justifiable residual covariances were coded such as correlated errors of manifest variables within the hypothesised factor. To visually simplify the path models, the coefficients of the relations between the factors are not presented but can be found in Appendix 20 (and Appendix 22 for number wizards).

4.2.1 Descriptive Statistics.

The descriptive analysis of all measures taken at all testing points can be seen in Table 4.1. Descriptive statistics were conducted using IBM SPSS Statistics 22. The standardized tests TROG-2 and BPVS-III were reported as raw scores, but standard scores of the sample were calculated and are given here. The sample was surprisingly lower than the population concerning grammatical ability (TROG-2, M = 91.54, SD = 1.57) but still within normal ranges. This is most likely due to the fact that 29 children were younger than four while standardization for TROG-2 starts at four. Vocabulary skills were representative of the population (BPVS-III, M = 101.58, SD = 1.57). Children's math-related language comprehension improved over time and ceiling effects were present at Time 5. It is worth mentioning that only their understanding of *more* was assessed at Time 1.

Children's performance on all measures of transcoding (number writing, number reading and number identification) increased over time although difficulty levels were adjusted to avoid ceiling effects. A clear ceiling effect was present for number reading at Time 1 with 45% of children achieving a maximum score. Furthermore, number reading and number writing were slightly negatively skewed at most testing points, contrary to number identification, with neither floor nor ceiling effects present. It is noteworthy that children's performance on the counting one-to-40 task at Time 3 was at ceiling level with more than half of the sample scoring at maximum. It seems that most children have mastered this section of the spoken count sequence at the age of five years and six months.

Children's central executive functioning at Times 4 and 5 was assessed using a visual search task. Each child's d prime was calculated as a measure of sensitivity, whereas proportion correct is affected by both sensitivity and bias (range is -8.6 to 8.6). At Time 4, the easy version was approaching ceiling level (*M* = 6.02, *SD* = 1.95 with 27 children achieving a perfect score). In contrast, the Time 4 hard version as well as both Time 5 hard versions show a more balanced range (means around the value four) indicating that the harder version is a more sensitive measure of central executive functioning at the age of five to six.

Focusing on the descriptive statistics of arithmetic, there was a ceiling effect at Time 1 with 17 children reaching the maximum score. Hence the arithmetic task was administered with a time constraint at Time 2 (children had three minutes to complete ten additions). The TOBANS was introduced at Time 3. All measures of TOBANS improved over time and there were floor effects present in the 'addition with carry' subtask at Times 3 and 4 and the 'subtraction' subtask at Time 4. The WIAT-II was assessed at Time 5. All items of the test were administered and converted into standard scores. This sample scored within normal range (*M* = 102.77, *SD* = 1.14) suggesting that the sample is representative of the population regarding mathematical skills. In subsequent analyses, the first six items (identifying and writing Arabic numerals) were excluded, in order to focus on arithmetic skills *per se*.

Table 4.1

Mean and standard deviations of predictor and criterion measures from all testing sessions

		Time 1	Time 2	Time 3	Time 4	Time 5
		M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Nonverbal IQ	Raven's CPM	6.45 (1.57)				
Working Memory	Visual Search	Easy: 3.98 (1.70) [3]*		Easy: 5.43 (1.35) [10] Hard: 4.16 (.37)	Easy: 6.021(.95) [27] Hard: 4.28 (.60)	A: 4.52 (.39) B: 4.57 (.73)
Language Comprehension	TROG-2	3.15 (2.63)*				
Vocabulary	BPVS-III	58.26 (16.77)*				
Literacy	TOWRE-2					Words: 44.04 (15.44) Nonwords: 24.81 (10.97)
	SWST					23.68 (6.06) [15]
Math-related Language	TRC					
	more	5.90 (2.00)*	7.35 (2.44) [3]	8.14 (2.67) [7]		10.19 (1.61) [15]
	less		5.32 (2.52) [2]	6.06 (2.67) [1]		9.11 (2.74) [14]
	overall		12.79 (4.20)*	15.51 (3.89)*		19.30 (3.71) [7]*

Numerical Knowledge	Number Writing	6.86 (5.79)*	7.93 (2.96) [15]*	9.24 (2.42) [1]	8.44 (3.03) [10]	13.07 (4.17) [12]
	Number Reading	8.02 (2.67) [45]*	14.76 (3.06) [4]*	12.77 (4.34) [6]*	15.47 (5.49) [7]*	23.95 (6.15) [14]*
	Number Identification	7.26 (2.60)*	10.99 (3.43) [1]*	11.99 (3.81)*	16.04 (4.24)*	22.81 (5.25) [1]*
	Rote Counting	14.78 (12.84)*	44.38 (30.23)*	1 to 40: 34.84 (8.51) [60] 94 to 110: 10.30 (5.34) [26] 25 back.: 7.01 (7.73) [12]*		
Magnitude Comparison	Digit Close	10.11 (3.21) [7]	11.97 (3.07) [20]	10.90 (4.92)	13.04 (4.46)	16.89 (4.57)
	Digit Far	10.73 (3.59) [7]	13.64 (3.33) [60]	14.32 (4.85)	17.20 (5.33)	22.32 (5.82) [2]
	NS FS Close	10.26 (2.16) [1]	10.64 (2.20) [1]	9.99 (4.14)	12.12 (4.07)	15.14 (4.33)
	NS FS Far	12.87 (2.57) [17]	14.13 (2.04) [34]	15.49 (6.11)	17.99 (5.31)	21.68 (6.32) [3]
	NS FS 2:3		13.18 (2.35) [24]	13.90 (5.12)	17.63 (5.98)	21.98 (6.41) [4]
	NS FS 3:4	10.76 (2.72) [4]	11.97 (2.44) [16]	12.82 (5.38)	15.56 (5.23)	19.68 (5.86) [2]
	NS FS 5:6	10.58 (2.29) [3]	11.23 (2.28) [3]	9.86 (4.95)	12.40 (3.94)	14.81 (4.95)
	NS SA Close	10.29 (2.28) [1]	10.67 (2.00)	8.99 (3.64)	10.57 (3.77)	12.47 (3.93)
	NS SA Far	13.24 (2.43) [23]	13.49 (2.09) [19]	14.48 (5.93)	17.78 (5.45)	21.96 (6.30) [2]
	NS SA 2:3	11.42 (2.42) [5]	12.37 (2.28) [8]	12.68 (5.51)	16.14 (5.42)	19.12 (6.45)
	NS SA 3:4	10.81 (2.03) [2]*	11.56 (2.23) [3]	10.92 (5.14)	13.40 (4.90)	17.35 (5.45)
	NS SA 5:6		10.45 (2.11) [1]*	9.60 (4.13)	11.34 (3.73)	13.23 (4.60)

Arithmetic	Addition Tasks	6.33 (3.21) [17]*	A: 5.23 (2.51) [3] B: 5.15 (2.55) [8]*			Addition: 6.02 (1.51) [10] Subtraction: 5.04 (2.41) [10]
	TOBANS					
	Addition			6.23 (4.55)	8.36 (5.09)	12.74 (8.66)
	Addition w/ carry			1.75 (2.20)	2.56 (2.74)	5.07 (5.01) [1]
	Subtraction				5.30 (4.12)	8.44 (5.10)
	Approximate Arithmetic			Symbolic: 15.16 (3.74) NS: 16.16 (3.51)*	Symbolic: 16.44 (3.67) [4] NS: 18.07 (3.03) [1]*	Symbolic: 18.53 (3.94) [13] NS: 19.72 (2.86) [5]*
	WIAT					4.00 (2.27)

*Notes. M = mean age. SD = standard deviation * individually administered tasks. The number of children scoring at maximum are shown in square brackets. All scores are presented as raw scores. For the Magnitude Comparison Tasks: NS = nonsymbolic. FS = fixed size trials. SA = surface-area matched trials.*

4.2.2. Structural Equation Modelling.

4.2.2.1 Concurrent prediction at Time 1.

The analysis of concurrent prediction of arithmetic at pre-school age included the latent independent predictors nonverbal intelligence (Raven's CPM), general language comprehension (BPVS-III and TROG-2), math-related language (TRC) and transcoding (number writing, reading and identification), counting skills (rote counting) as well as the two magnitude comparison constructs (symbolic and nonsymbolic comparison) and the dependent outcome variable arithmetic (data of addition task was split into two manifest variables – odd and even numbered problems). Nonverbal intelligence, math-related language comprehension and counting were each assessed by only one indicator (Raven's CPM, TRC and rote counting), which may distort the data as a result of measurement errors. Thus, these indicators were pre-specified with an error reflecting the reliability of the variable calculated on the sample.

All manifest variables loaded significantly on their proposed latent constructs. It is worth mentioning that the easy nonsymbolic comparison *surface-area matched* ratio 2:3 has the weakest loading of the comparison tasks. The path model depicted in Figure 4.3 provided an excellent fit, $\chi 2$ (84) = 92.439, p = .248, *RMSEA* = .032 (90% CI = .000 - .065), *CFI* = .981, *SRMR* = .059. The latent variables nonverbal intelligence and transcoding were the only unique predictors of children's performance on simple arithmetic task at Time 1 (70.2% of variance was explained). This result suggests that only nonverbal intelligence and transcoding, children's ability to translate between verbal number codes and Arabic numerals, may be crucial to the development of arithmetic at the age of 4 years 2 months, before the beginning of formal education.

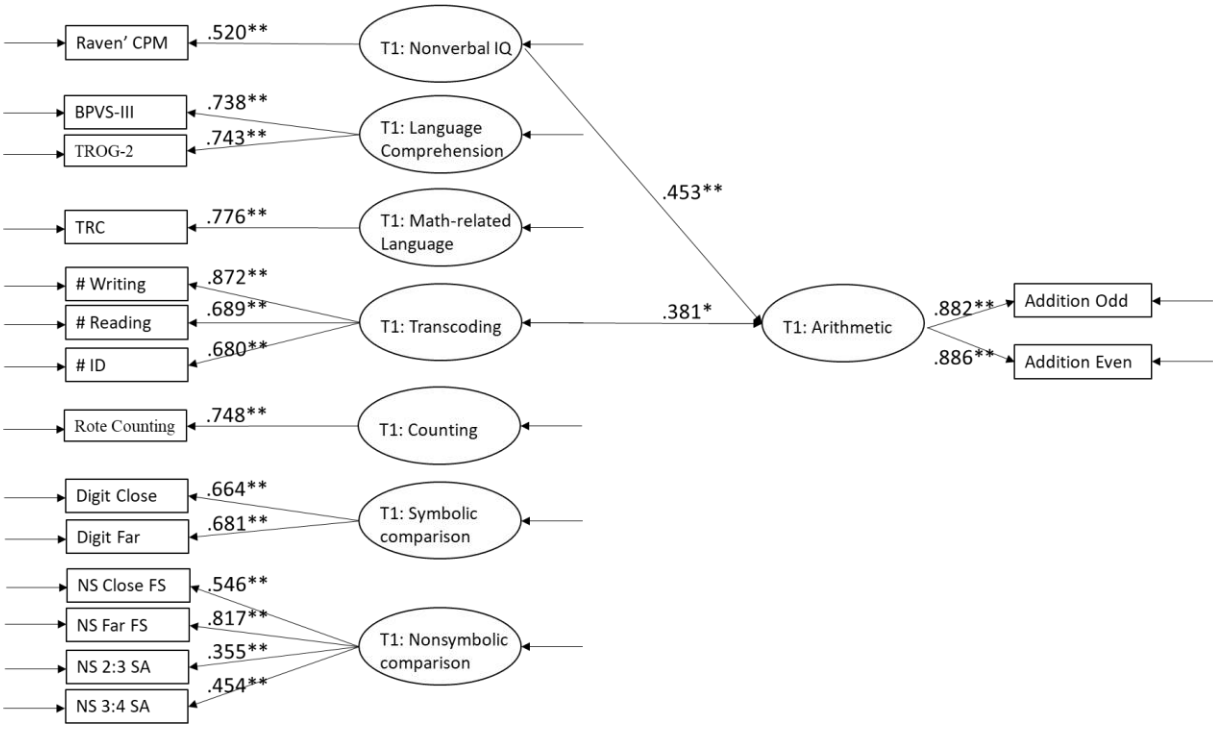


Figure 4.3. Concurrent associations of arithmetic assessed at Time 1. * p < .05. ** p < .01.

Similar to the previous chapter, the same Time 1 model as used above was re-run only using the data for *number wizards* (45 children that achieved the maximum score on the numeral reading task). Based on the findings from chapter three, the magnitude comparison tasks formed on factor rather than two distinct factors. Also, *surface-area matched* stimuli were removed due to the fact that they did not significantly load onto the hypothesised general magnitude comparison factor and the number reading task was removed because the analysis only investigated children who have achieved the maximum. The model, shown in Figure 4.4, provided an excellent fit to the data, $\chi 2$ (51) = 46.587, *p* = .649, *RMSEA* = .000 (90% CI = .000 - .081), *CFI* = 1.00, *SRMR* = .058. Similar to the above model, transcoding was strongly predicting children's early arithmetic scores. Surprisingly, nonverbal intelligence was not a unique predictor as seen in the model containing data from the whole sample. Yet math-related language comprehension was the second unique predictor of performance on arithmetic tasks at four years of age, with transcoding being the stronger predictor. This model explained 54.8% of variance.

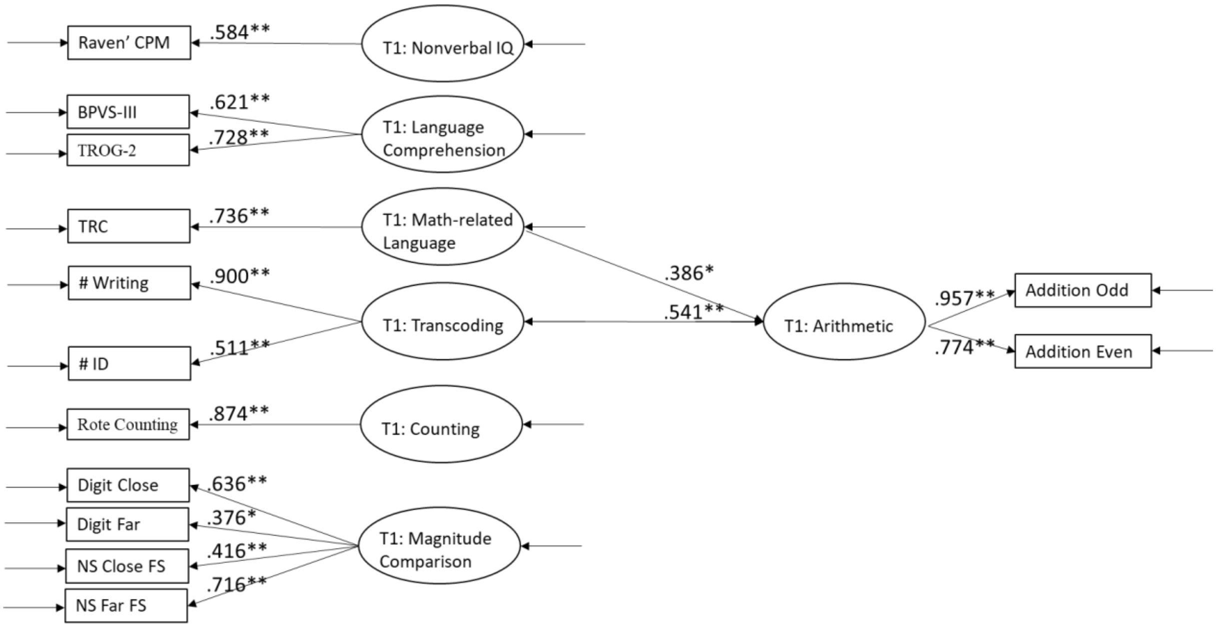


Figure 4.4. Number wizards' concurrent associations of arithmetic assessed at Time 1. * p < .05. ** p < .01.

Correlations. The correlations, based on the full sample, between the latent constructs are shown in Table 4.2. The latent outcome variable arithmetic correlated with all other variables, but the highest correlation was with nonverbal intelligence and transcoding which confirms the findings of prediction of the SEM path model. Furthermore, both symbolic and nonsymbolic magnitude comparison were also strongly associated with arithmetic.

Interestingly, general language comprehension was strongly related to math-related language comprehension and symbolic magnitude comparison. The former is easy to explain; both tests measure a form of language comprehension. Additionally, the TROG included a section which directly tested children's understanding of more and less which was the focus of the math-related language comprehension task. The latter association with symbolic magnitude comparison may suggest that, at this age, children highly rely on language skills to compare two numbers due to the fact that they have not completely mastered the Arabic numeral system. This relationship may fade over time when children grow more confident in working with numerals.

Table 4.2

Correlations between the predictor measures and the criterion measures at Time 1 (n = 100)

	1	2	3	4	5	6	7	8
1. Nonverbal Intelligence	---	.451**	.439**	.409	.229	.437*	.632**	.757**
2. Language Comprehension		----	.621**	.529**	.290	.605**	.469**	.477**
3. Math-related Language			---	.346*	.212	.511**	.376**	.398**
4. Transcoding				----	.666**	.672**	.415**	.638**
5. Counting					----	.406*	.332*	.399**
6. Symbolic Comparison						----	.556**	.525**
7. Nonsymbolic Comparison							----	.540**
8. Arithmetic								----

Notes. Pearson product-moment correlation coefficient. * $p < .05$. ** $p < .01$.

The variable transcoding correlated highly with counting as well as symbolic comparison whereas the relation to nonsymbolic comparison was weaker. It is worth mentioning that symbolic and nonsymbolic comparison were only moderately correlated which supports the idea of two individual constructs at this age.

4.2.2.2 Concurrent prediction at Time 2.

The path model investigated prediction of arithmetic in reception class, the first stage of formal schooling, between the independent variables math-related language (*more* and *less*), transcoding (number writing, reading and identification), counting skills (rote counting) as well as the two magnitude comparison constructs (symbolic and nonsymbolic comparison) and the dependent variable arithmetic (addition task, form A and B). Because counting was assessed by one indicator, it was pre-specified with an error to avoid distortions caused by measurement errors.

Figure 4.5 shows the unique predictors of arithmetic at reception class age (4 years and 11 months) confirming that the manifest variables load onto their hypothesised latent factors. Surprisingly, the prediction of transcoding was not significant for this age group. However, the model found that math-related language comprehension and counting were the only unique predictors of arithmetic with an excellent model-fit to the data, $\chi 2$ (66) = 63.676, p = .558, *RMSEA* = .000 (90% CI = .000 - .051), *CFI* = 1.00, *SRMR* = .046. The predictors explained 68.2% of variance. It is worth mentioning that not all tasks of Time 1 were retested and the list of tests was limited to mainly tasks strongly connected to numeracy skills. Measures of nonverbal intelligence and general language comprehension were not assessed.

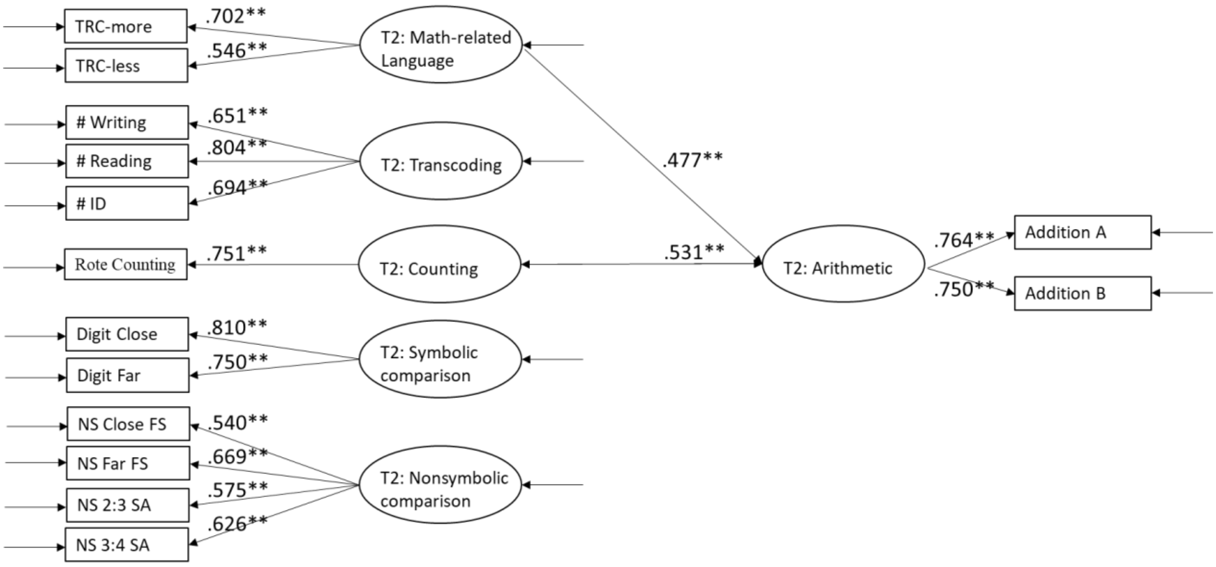


Figure 4.5. Concurrent associations of arithmetic assessed at Time 2. * p < .05. ** p < .01.

Correlations. As expected, all latent factors correlated with arithmetic (correlations are shown in Table 4.3). However, the strongest correlations at Time 2 were counting and transcoding and math-related language. Interestingly, the correlation with transcoding was higher than math-related language although transcoding did not uniquely predict arithmetic.

Further inspection of the correlation matrix revealed a strong association between transcoding and counting skills (r = .774), suggesting that both factors share similar cognitive resources and constructs. The factor math-related language highly correlated with all other factors but counting. The strongest relation was with nonsymbolic magnitude comparison which may due to the fact that the nonsymbolic comparison greatly draws on children's understanding of the term *more* ('Which box has more dots?') to excel on the task.

Table 4.3

Correlations between the predictor and the criterion measures at Time 2 (n = 117)

	1	2	3	4	5	6
1. Math-related Language	---	.554**	.342	.582**	.637**	.658**
2. Transcoding		----	.774**	.735**	.495**	.675**
3. Counting			----	.622**	.549**	.693**
4. Symbolic Comparison				----	.603**	.608**
5. Nonsymbolic Comparison					----	.595**
6. Arithmetic						----

Notes. Pearson product-moment correlation coefficient. * $p < .05$. ** $p < .01$

4.2.2.3 Concurrent prediction at Time 3.

Latent independent factors at Time 3 included transcoding, counting (counting to 40, counting from 94 to 110 and counting backwards from 25), as well as the two magnitude comparison constructs (symbolic and nonsymbolic comparison) and the dependent variable arithmetic (TOBANS, addition and addition-with-carry).

The path model predicting arithmetic in the autumn term of Year One (Figure 4.6) provided an acceptable fit to the data, $\chi2$ (66) = 86.382, p = .047, *RMSEA* = .052 (90% CI = .000 - .080), *CFI* = .978, *SRMR* = .049 (loadings of manifest variables onto hypothesised latent factors was reasonable). Surprisingly, transcoding was not significantly predicting arithmetic and, interestingly, only symbolic magnitude comparison and counting were unique predictors of arithmetic (86.5% of variance was explained).

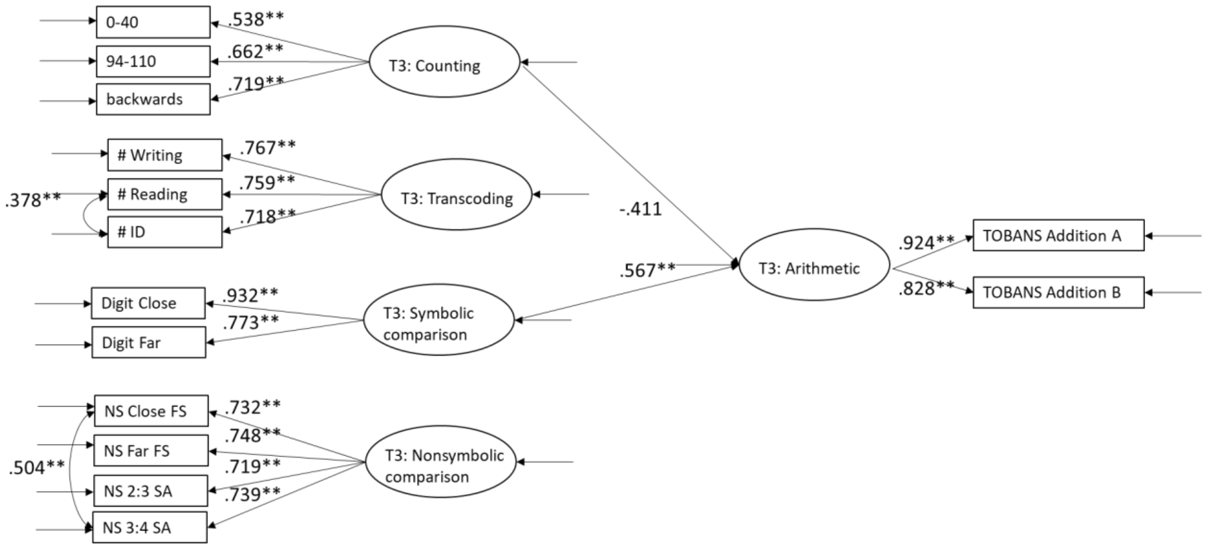


Figure 4.6. Concurrent associations of arithmetic assessed at Time 3. * p < .05. ** p < .01.

Correlations. The correlations between the latent constructs are shown in Table 4.4 and showed that all latent factors were related to arithmetic. Transcoding and both magnitude comparison factors showed the strongest relation.

Table 4.4

Correlations between the predictor and the criterion measures at Time 3 (n = 116)

	1	2	3	4	5
1. Transcoding	----	.880**	.794**	.636**	.785**
2. Counting		----	.665**	.587**	.863**
3. Symbolic Comparison			----	.869 **	.805**
4. Nonsymbolic Comparison				----	.729**
5. Arithmetic					----

*Notes. Pearson product-moment correlation coefficient. * p < .05. ** p < .01*

It is worth mentioning that symbolic magnitude comparison was highly related to transcoding and even more highly to nonsymbolic comparison. The latter supports the finding in Chapter 3 that the relationship between symbolic and nonsymbolic magnitude comparison slowly shifts around the time children enter formal schooling, moving from a two-factor structure towards one general magnitude comparison factor. Transcoding was highly correlated to arithmetic and counting. Interestingly, the path model found a nonsignificant prediction between arithmetic and transcoding though both correlated highly.

4.2.2.4 Concurrent prediction at Time 4.

The concurrent relationship model at Time 4 investigated prediction of arithmetic in the spring term of Year One. At this time all participants had been in formal education for at least two terms. The model comprised of the dependent variable arithmetic (TOBANS, addition, addition-with-carry and subtraction) and the latent independent variables transcoding (number writing, reading and identification), executive functioning and general magnitude comparison (symbolic and nonsymbolic comparison).

The SEM path model of prediction of arithmetic (Figure 4.7) showed consistent high loading of the manifest variables onto their hypothesised latent factors. The model provided an acceptable fit to the data, $\chi 2$ (72) = 88.563, p = .09, *RMSEA* = .062 (90% CI = .000 - .102), *CFI* = .961, *SRMR* = .066. Interestingly, the only unique predictor in this model was the general magnitude comparison variable, nonetheless transcoding was marginally significantly predicting early arithmetic. It is worth mentioning that the number of independent variables was very limited due to testing time constraints. Only 54% of variance is explained by the model.

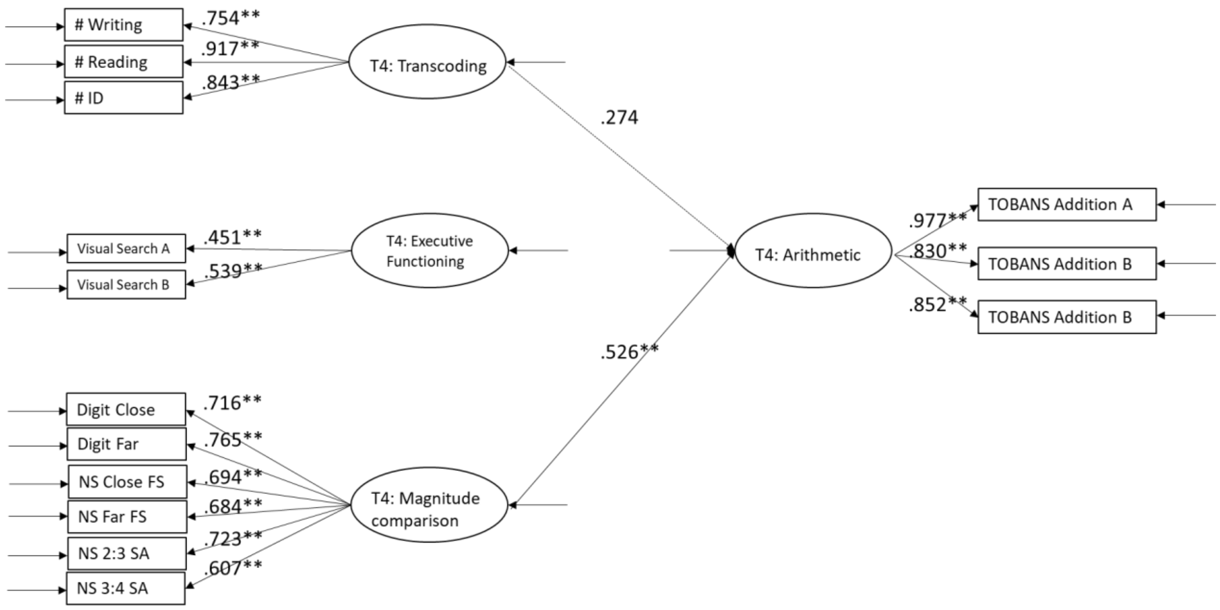


Figure 4.7. Concurrent associations of arithmetic assessed at Time 4. * p < .05. ** p < .01.

Correlations. In contrast to previous testing sessions, the strongest association with arithmetic was found with the unique predictor general magnitude comparison (correlations are shown in Table 4.5). The correlations illustrate that general magnitude comparison tasks are highly correlated, not surprisingly, with both transcoding and executive functioning.

Table 4.5

Correlations between the predictor and the criterion measures at Time 4 (n = 115)

	1	2	3	4
1. Transcoding	---	.658**	.528*	.619**
2. Magnitude Comparison		----	.831**	.705**
3. Executive Functioning			----	.581**
4. Arithmetic				----

*Notes. Pearson product-moment correlation coefficient. * p < .05. ** p < .01*

4.2.2.5 Concurrent prediction at Time 5.

The last path model (Figure 4.8) investigated concurrent prediction of arithmetic at Time 5 (summer term of Year One) between the predictor variables central executive functioning (Visual Search A and B), literacy (TOWRE-2 nonwords, SWST), transcoding (number writing, reading and identification) and, based on previous results (see Chapter 3) one general magnitude comparison construct and the outcome variable arithmetic (TOBANS, addition, addition-with-carry and subtraction). The following correlated errors were included: Number writing with SWST because both task require the children to write the answer and may presumably share similar writing processes, and fixed size close items with surface-area matched ratio 3:4 items as they both share the same methodology and both are the harder version of each size condition.

The manifest variables load satisfactorily onto their latent factors. As shown in previous testing time points, the latent variable transcoding and the general magnitude comparison factor were the only unique predictors of arithmetic of six year-old children. Although the chi-squared difference test was significant indicating that the theoretical model may be different from the observed data, all other indices of goodness of fit were acceptable, $\chi 2$ (94) = 144.445, p = .002, *RMSEA* = .064 (90% CI = .041 - .086), *CFI* = .964, *SRMR* = .052. This model only explained 45.5% of the variance indicating that there may be more predictors that contribute to the prediction of arithmetic scores at the end of Year One. Children's understanding of Arabic numerals was a slightly stronger predictor of arithmetic scores than general magnitude comparison.

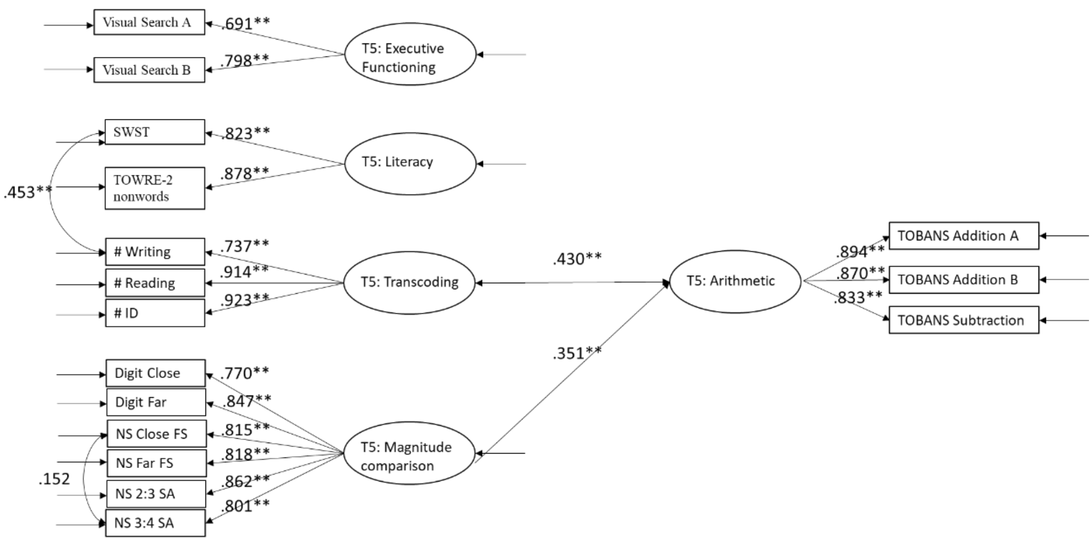


Figure 4.8. Concurrent associations of arithmetic assessed at Time 5. * p < .05. ** p < .01.

Correlations. The correlation matrix for Time 5 latent factors is shown in Table 4.6. All independent factors correlated highly with arithmetic assessed at Time 5 with the highest correlations with transcoding and magnitude comparison. Besides, central executive functioning was related with magnitude comparison and, of particular interest, literacy with transcoding. The latter correlation may be due to shared linguistic components that influence both factors.

Table 4.6

Correlations between the predictor and the criterion measures at Time 5 (n = 119)

	1	2	3	4	5
1. Executive Functioning	---	.469**	.521**	.693**	.467**
2. Literacy		----	.724**	.545**	.501**
3. Transcoding			----	.493**	.600**
4. Magnitude Comparison				----	.563**
5. Arithmetic					----

*Notes. Pearson product-moment correlation coefficient. * p < .05. ** p < .01.*

4.3 Conclusion.

The scope of this chapter was to investigate concurrent predictors of arithmetic skills at five time points from pre-school through to the end of the first year of formal schooling, taking snapshots of development at different stages. There are limitations to the conclusions that can be drawn, since the measurements taken differ at different time points. Nonetheless, the results are informative concerning the process of development, and provide a useful background to the research exploring longitudinal prediction of arithmetic.

To sum up the findings, nonverbal intelligence and transcoding were the only unique predictors of children’s performance on arithmetic tasks at age of four years and three months (Time 1). Although previous findings showed that general intelligence affects children’s early arithmetic skills were replicated (Cowan et al., 2005; Noël, 2009) with children’s nonverbal intelligence being a slightly stronger predictor of variance in arithmetic tasks, neither of these studies assessed children’s numeracy skills measured by transcoding. Göbel et al. (2014) showed that transcoding plays a crucial role in the development of arithmetic in six year-old children. The current study transferred these findings to pre-school children showing that four year-olds ability to translate between the Arabic numerals and verbal codes also affects arithmetic performance. However, a different pattern emerged when investigating the performance of number wizards (high achievers on the number reading task) on arithmetic. Transcoding was the strongest unique predictor of arithmetic scores. Surprisingly, nonverbal intelligence was not a unique predictor, yet math-related language comprehension was uniquely predicting performance on arithmetic tasks. It seems that either nonverbal intelligence may not be as important once children’s number recognition (number reading) was taken into account or that the relationship in number wizards is different from the general population. This could be the subject of further study. However, the findings support the idea that knowing your numbers is crucial for the development of early arithmetic and suggest that children who have already mastered the Arabic numerals from one to ten may not rely on cognitive processes such as nonverbal intelligence but rather specialised math-related skills such as math-related language and transcoding (numerical knowledge).

A different pattern of relationship can be found at Time 2, with math-related language comprehension and counting being the only two unique predictors of children’s performance on arithmetic tasks. It seems that counting may be the stronger predictor confirming previous research findings that counting may be crucial for attainment of arithmetic (Butterworth, 2005; Desoete and Grégoire, 2006; Nunes and Bryant, 1996; Gelman and Gallistel, 1978). Indeed, researchers reported that counting is important for calculation (Ansari, Donlan, Thomas, Ewing, Peen and Karmiloff-Smith, 2003; Cowan et al, 2005). Furthermore, Donlan et al. (2007) found a strong association between counting and calculation suggesting that the performance on both tasks draw from a common representational system.

Moreover, the results regarding math-related language comprehension support the notion that language impacts early arithmetic (Donlan et al., 1998; Donlan, and Gourlay, 1999; Fazio, 1994; 1996; Donlan et al., 2007; Cowan et al., 2005; Kleemans et al., 2011; 2012). However, most studies more general language skills neglecting language specific to mathematics. Indeed, this study found that arithmetic may be predicted by language specific to mathematics.

Neither nonverbal intelligence nor transcoding significantly contribute to explaining the variance in children’s arithmetic scores at 4;11 years. However, the Time 2 path model only included few variables, most of them numeracy tasks. It will be interesting to examine if counting and math-related language may longitudinally become a key foundation of arithmetic.

At Time 3, counting was a unique predictor of arithmetic, as was symbolic comparison tasks. These results confirm previous findings that counting (Butterworth, 2005; Desoete and Grégoire, 2006; Nunes and Bryant, 1996; Gelman and Gallistel, 1978; Ansari et al., 2003; Cowan et al, 2005) and symbolic magnitude comparison (Holloway and Ansari, 2009; de Smedt et al., 2013; Siegler, 2016) are important concurrent predictors of children's arithmetic ability.

Comparable to the results from Time 3, magnitude comparison seems to play a crucial part in children's performance on arithmetic tasks at Time 4. The contribution of transcoding was only marginally significant which suggests that it may not substantially contribute to explaining the variance of arithmetic at five years and ten months of age. It seems that the influence of transcoding diminishes over time in favour of the strengthened relation between arithmetic and magnitude comparison. Also, this is the first time that symbolic and nonsymbolic comparison tasks load onto a unitary factor rather than two distinct factors which may explain why magnitude comparison impacts arithmetic so strongly.

At Time 5, transcoding and the general magnitude comparison factor uniquely predicted arithmetic scores, with transcoding being the stronger predictor. These findings support the hypothesis that the ability to translate between Arabic numeral and their verbal code crucially impact the development of early arithmetic. Interestingly, it is widely held that working memory contributes to arithmetic skills in typically developing children (Berg, 2008; Kleemans et al., 2012) and central executive functioning in particular (Gathercole and Pickering, 2000). Gilmore et al. (2014) proposed that executive functioning may consist of three types: working memory, inhibition and shifting. Previous research typically used recall tasks, such as digit recall, listening recall or backward digit recall which are more in line with monitoring and manipulating information (working memory). Also, digit recall tasks may share cognitive processes with arithmetic because they tap into numerical knowledge and processes. The executive functioning task in this study was assessing inhibition aspects of executive functioning which may explain why the current measure of executive functioning was not a powerful predictor of early arithmetic at Times 4 and 5.

Overall, transcoding, children's ability to translate between spoken and symbolic form of numbers, seems to play the most consistently important role in the development of early arithmetic skills. Transcoding may not have been the strongest or only predictor at times, and other factors may also impact the development of early arithmetic at different time points. At an early, pre-school age, it appears that nonverbal intelligence, counting and math-related language, particularly children's understanding of more, in addition to transcoding, affect the performance on arithmetic tasks. These relations however, weaken in favour of the relationship with magnitude comparison in early school years (Year One). After children entered the formal schooling system, both transcoding and a general magnitude comparison factor were crucial for arithmetic development.

The transcoding factor entails both Arabic-digit knowledge and place-value understanding. According to previous research, children's understanding of place-value may be a key foundation for the development of later arithmetic skills. Möller, Pixner, Zuber, Kaufmann, and Nürk (2011) showed that seven-year-olds place-value understanding predicted their performance on addition tasks two years later.

Additionally, this study confirms previous research findings which suggest that Arabic-digit knowledge at school entry may play a crucial role on children's arithmetic development (Kolkman et al., 2013; Krajewski and Schneider, 2009; Mundy and Gilmore, 2009). This relationship appears to be directly analogous to the critical longitudinal role of early letter knowledge on the development of reading skills (Caravolas et al., 2012; Hulme, Bowyer-Crane, Carroll, Duff, and Snowling, 2012). Indications from latent factor correlations at Time 5 suggest that learning arithmetic may share some developmental pathways with learning to read. It appears that learning the symbol set (Arabic numerals or letters) and their verbal labels is a critical foundational skill for later literacy and arithmetic skills.

It must be critically mentioned that not all measures were assessed at all testing points, thus constraining conclusions drawn about the concurrent prediction of arithmetic and the change of the relationships with arithmetic over time. Also, the testing procedure of tasks was changed to adjust for children's growing learning experience in the tasks measured (see assessment of magnitude comparison and arithmetic for more details). Further studies are needed to investigate the concurrent prediction using the same tasks at all testing points to enable comprehensive conclusions about prediction of early arithmetic and how the concurrent relationships may change over time.

Although cross-sectional relationships may draw attention towards special and changing relations between arithmetic and its precursors, they do not test longitudinal prediction of arithmetic skills.

Chapter 5. Longitudinal Prediction of Early Arithmetic

Recent longitudinal studies investigating the influence of ANS measures on math achievement produced mixed results (Desoete et al., 2012; Lyons et al., 2014; Kolkman et al., 2012). Lyons and colleagues (2014) explored the prediction of arithmetic through primary school and found no evidence that individual differences on nonsymbolic comparison were a unique predictor of arithmetic scores at any grade. Kolkman et al. (2012) examined the relationship between arithmetic and nonsymbolic, symbolic and number estimation skills at age four, five and six. The findings suggest that nonsymbolic, symbolic and number estimation skills were separate skills at a younger age integrating over time into one general numeracy skills concept. Only children's number estimation skills were uniquely predictive of math performance at six years. A recent study by Göbel et al. (2014) addressed the relation between nonsymbolic and symbolic judgement tasks and their role as longitudinal predictors of arithmetic development in six-year-olds. The path model revealed that number identification was the most powerful longitudinal predictor of arithmetic skills at age seven, apart from the auto-correlate.

The main focus of this chapter is to identify the longitudinal predictors of arithmetic across a two year period focusing on the role of ANS and language in particular. To capture children's pre-school abilities before formal schooling, it was decided that the base model should be the concurrent model at Time 1 (Chapter 4). This base model was then regressed onto arithmetic performance at Times 2, 3, 4 and 5.

5.1 Methods.

The same participants were used as described in Chapter 2 (p. 42)

5.1.2 Materials.

Children were assessed on the following measures.

5.1.2.1 Baseline Prediction Model assessed at Time 1.

The following tasks were administered individually to the four nursery classes in the summer term of the nursery age (four years; see Chapter 4, pp. 76-79 for more details): Nonverbal intelligence (Raven's CPM; Raven et al., (1993), grammatical ability (TROG-2; Bishop, 2003), vocabulary (BPVS - III; Dunn et al., 2010), specific math-related language ability (TRC), transcoding (Number Identification, Number Writing and Reading Arabic numerals), rote counting and symbolic and nonsymbolic magnitude comparison.

5.1.2.2 Arithmetic Skills.

Time 1. The children's basic arithmetic skills were assessed using simple addition problems. The test comprised ten simple additions with sums less than ten (1 + 3; 2 + 1; 2 + 2; 1 + 4; 3 + 1; 1 + 5; 2 + 3; 1 + 6; 3 + 3; 4 + 4). All arithmetic problems were presented in Arabic notation (MS Office 2013, Comic Sans MS, size 260) and, simultaneously, in spoken form most familiar to the child. Problems were arranged so that additions with same sums or similar summands were never adjacent. Children were encouraged to use wooden sticks provided or their fingers if needed. Before two practice problems (1 + 1, 1 + 2) were administered, the preferred method of referring to additions ("add" or "plus") was determined by asking the teachers. Testing was only terminated early if a child showed signs of confusion or lack of concentration. The maximum score was ten.

Time 2. Due to the ceiling level performance at Time 1, the following adjustments have been made to the basic addition task:

Two parallel forms of the tasks have been created which comprised of ten simple additions with sums less than ten (both forms were equal in difficulty level; Form A: 1 + 3; 2 + 1; 1 + 5; 2 + 3; 4 + 5; 7 + 2; 3 + 5; 4 + 2; 5 + 2 and 2 + 6; Form B: 1 + 4; 3 + 1; 2 + 5; 4 + 2; 1 + 6; 3 + 6; 2 + 7; 6 + 2; 4 + 3 and 3 + 5). To raise the sensitivity of the task further, children were given only three minutes to solve as many problems as possible. The two forms were given in two separate testing sessions. To avoid training effects, the order of the forms was counterbalanced. The total number of correctly solved problems was recorded.

Time 3. *Fluency.* Children's speeded arithmetic skills (fluency) was assessed using the 'addition' and 'addition with carry' subtests of the TOBANS (Brigstocke et al., 2016). Children were asked to complete as many arithmetic problems as possible in one minute. In the 'addition' subtask, children were presented with simple addition problems with sums less than ten and in the 'addition with carry' subtask the sums were greater than ten but less than twenty. One point was awarded even if the numeral was written backwards (maximum score $_{addition}$ = 90; maximum score $_{addition\ with\ carry}$ = 30). This task was administered as a group task according to the manual.

Time 4. *Fluency.* In addition to the same tasks as at Time 3 (addition and addition with carry), children were also presented with the 'subtraction' subtask (Brigstocke et al., 2016). Similar to addition, children were asked to solve as many of the 90 subtraction problems as possible in one minute.

Time 5. *Fluency.* The same tasks as at Time 4 were used (see above for more details).

Accuracy. Children's basic arithmetic skills were assessed using the Numerical Operations subtest of the second edition of the Wechsler Individual Achievement Test (WIAT-II; Wechsler, 2005). The first six items (identifying and writing Arabic numerals) were excluded because we were only interested in a more conventional measure of arithmetic. The test was executed according to the manual and children were allowed to complete the task in their own time (maximum score = 25).

5.1.3. Procedure.

All measures in this chapter were part of comprehensive test battery. Tests were divided into 20 to 40-minute-blocks at each time point to counterbalance effects of order such as learning and motivational effects. Even the testing order within each block was counterbalanced. Testing was carried out five times over a 25-month period from the summer term of nursery through to the summer term of Year One. Wherever possible, each child was seen by the same experimenter. The main researcher was assisted by several research assistants from undergraduate psychology classes. They were trained on how to administer the test battery and were given instructions on how to work with young children. The baseline predicting model was assessed individually at Time 1. Children's arithmetic skills were tested individually at Times 1 and 2. Arithmetic tasks at Times 3, 4 and 5 were administered in small groups in a separate room or another quiet place in the school. Each child met with the experimenter ideally two to four days in a row, depending on the number of blocks, to avoid lack of motivation or concentration. If testing in groups, the ratio of experimenters to children was 1:3.

Preliminary to the testing, the experimenters attended at least one day in each class so that the children got to know them and felt more comfortable around them. Moreover, the experimenters told each child that they would play games and asked questions such as “How are you?” or “How old are you?”.

All unstandardized tests included practice items. Concerning feedback, children received only concrete feedback on their performance for practice items and general praise and encouragement throughout the tests.

5.2 Results.

To answer this research question concerning the longitudinal prediction of early arithmetic, a set of SEM path models were estimated with Mplus Version 7 (Muthén and Muthén, 2013) using the Time 1 concurrent prediction model as the base model onto which arithmetic scores at later time points were regressed. Dependent variable was the latent arithmetic factor at each testing point and independent variables included the latent factors of the baseline Time 1 model: nonverbal ability, general language comprehension, math-related language comprehension, counting, transcoding, nonsymbolic and symbolic magnitude comparison as well as the auto-correlate arithmetic Time 2. To visually simplify the path models, the coefficients of the relations between the factors are not presented but can be found in Appendix 21 (and Appendix 22 for number wizards).

5.2.1 Descriptive Statistics.

The descriptive analysis of all measures taken at all testing points can be seen in Table 5.1. Descriptive statistics were conducted using IBM SPSS Statistics 22. It emerged that performance of all tasks assessed increased over time. Children’s math-related language comprehension improved over time, but clear ceiling effects were present at Time 5. It is worth mentioning that only their understanding of *more* was assessed at Time 1.

Children’s performance on number writing and number reading showed clear ceiling effects with 45% of children achieving a maximum score on number reading at Time 1. Furthermore, neither floor nor ceiling effects were present regarding the number identification task.

Focusing on the descriptive statistics of arithmetic, there was a ceiling effect at Time 1 with 17 children reaching the maximum score. Hence the arithmetic task was administered with a time constraint at Time 2 (children had three minutes to complete ten additions). Although fewer children scored at ceiling level at Time 2, the data suggested that this measure was still too easy at the age of four years and four months and would have probably been too easy at later testing points. Thus the TOBANS was introduced at Time 3. All measures of TOBANS improved over time and there were floor effects present at ‘addition with carry’ subtask at Times 3 and 4 and the ‘subtraction’ subtask at Time 4. Furthermore, the WIAT-II was assessed at Time 5. All items of the test were administered and converted into standard scores. This sample scored within normal range (M = 102.77, SD = 1.14) suggesting that the sample was representative of the population regarding mathematical skills. Raw scores were used in further analyses. However, the first six items (identifying and writing Arabic numerals) were excluded in order to avoid confounding with other measures, and to provide a purer and more conventional measure of arithmetic.

Table 5.1

Mean and standard deviations of predictor and criterion measures from all testing sessions

		Time 1	Time 2	Time 3	Time 4	Time 5
		M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Nonverbal IQ	Raven's CPM	6.45 (1.57)				
Working Memory	Visual Search	Easy: 3.98 (1.70) [3]*				
Language Comprehension	TROG-2	3.15 (2.63)*				
Vocabulary	BPVS-III	58.26 (16.77)*				
Math-related Language	TRC	5.90 (2.00)*				
Numerical Knowledge	Number Writing	6.86 (5.79)*				
	Number Reading	8.02 (2.67) [45]*				
	Number Identification	7.26 (2.60)*				
	Rote Counting	14.78 (12.84)*				

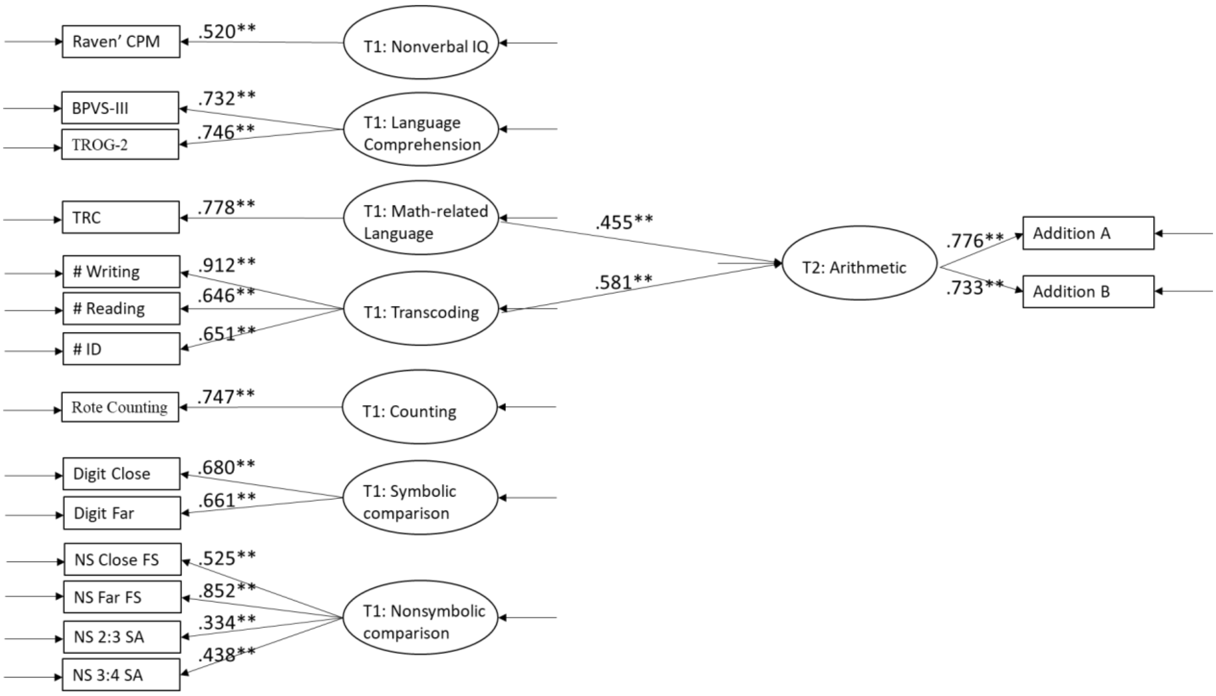
Magnitude Comparison	Congruent (FS)	33.89 (5.87)	61.15 (7.85)	57.74 (21.54)	75.32 (20.19)	84.92 (25.42)
	Incongruent (SA)	34.34 (4.98)	58.54 (7.31)	53.21 (20.20)	68.83 (18.69)	77.45 (23.11)
Arithmetic	Addition Tasks	6.33 (3.21) [17]*	A: 5.23 (2.51) [3] B: 5.15 (2.55) [8]*			Addition: 6.02 (1.51) [10] Subtraction: 5.04 (2.41) [10]
	TOBANS					
	Addition			6.23 (4.55)	8.36 (5.09)	12.74 (8.66)
	Addition w/ carry			1.75 (2.20)	2.56 (2.74)	5.07 (5.01) [1]
	Subtraction				5.30 (4.12)	8.44 (5.10)
	WIAT					4.00 (2.27)

Notes. *M = mean age. SD = standard deviation * individually administered tasks. The number of children scoring at maximum are shown in square brackets. All scores are presented as raw scores. For the Magnitude Comparison Tasks: NS = nonsymbolic. FS = fixed size trials. SA = surface-area matched trials.*

5.2.2 Predicting Time 2.

Investigation of longitudinal prediction of early arithmetic after nine months included the Time 1 base model (independent variables) regressed onto arithmetic (dependent variable) taken at Time 2 (mean age 4 years and 11 months). The Time 1 base model was established as follows: The analysis of concurrent prediction of arithmetic at pre-school age included the latent predictive variables nonverbal intelligence (Raven's CPM), general language comprehension (BPVS-III and TROG-2), math-related language (TRC) and transcoding (number writing, reading and identification), counting skills (rote counting) as well as the two magnitude comparison constructs (symbolic and nonsymbolic comparison) and the outcome variable arithmetic (data of addition task was split into two manifest variables – odd and even numbered problems). Nonverbal intelligence, math-related language comprehension and counting were each assessed by only one indicator (Raven's CPM, TRC and rote counting), which may distort the data as a result of measurement errors. Thus, these indicators were pre-specified with an error reflecting the reliability of the variable calculated on the sample. All manifest variables loaded significantly on their proposed latent constructs. It is worth mentioning that the easy nonsymbolic comparison *surface-area matched* ratio 2:3 had the weakest loading of the comparison tasks. The path model depicted in Figure 4.3 provided an excellent fit (see Chapter 4, p. 91).

The path model depicted in Figure 5.1 shows the longitudinal predictors of early arithmetic at Time 2. The model fit was acceptable, $\chi 2$ (84) = 99.058, p = .125, *RMSEA* = .036 (90% CI = .000 - .061), *CFI* = .962, *SRMR* = .064. Transcoding and math-related language comprehension uniquely predicted arithmetic at Time 2 (71.4% of variance explained). However, this models did not include the autoregressor (arithmetic taken at Time 1). The model with autoregressor arithmetic Time 1 (shown in Figure 5.2) provided a similar fit to the data as the previous model, $\chi 2$ (111) = 127.824, p = .131, *RMSEA* = .033 (90% CI = .000 - .056), *CFI* = .969, *SRMR* = .060. In Chapter 4 it was noted that nonverbal intelligence and transcoding uniquely contributed to explaining the variance of the concurrent outcome arithmetic at Time 1 (63.8% of variance explained), whereas math-related language comprehension just did not contribute to predicting arithmetic at Time 1. Thus the concurrent model was not confirmed by the prediction pattern for the longitudinal data (Times 1 and 2), with or without the autoregressor, in which transcoding and math-related language comprehension were significant unique predictors of arithmetic (72.9% of variance explained). Interestingly, the autoregressor did not



uniquely predict arithmetic at Time 2.

Figure 5.1. Prediction of arithmetic at Time 2 by Time 1 base model without autoregressor. * p < .05. ** p < .01.

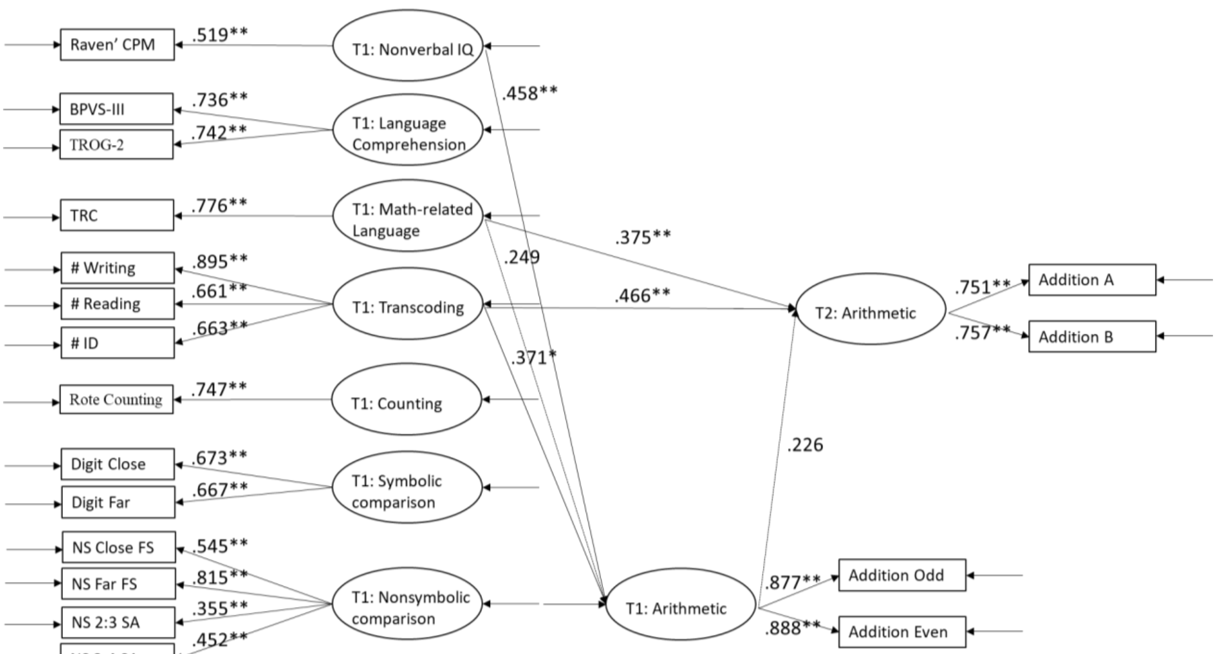


Figure 5.2. Prediction of arithmetic at Time 2 by Time 1 base model with autoregressor * p < .05. ** p < .01.

5.2.3 Predicting Time 3.

Next, longitudinal prediction of children's performance of arithmetic 16 months later (autumn term Year One) was modelled using the Time 1 base model (independent variables) and arithmetic outcomes of the TOBANS subtasks from Time 3 (independent variable). Because the path models with autoregressor (Times 1 and 2) were similar in fit and predictors, it was decided that only one path model with autoregressor will be compared henceforward to the model without the autoregressor. The Time 2 autoregressor was chosen based on the fact that assessment of arithmetic at Time 1 showed ceiling effects and differed in methodology (not constrained for time). First, the base model was regressed on Time 3 arithmetic without the autoregressor and the second model investigated prediction with Time 2 autoregressor present. The models included correlated error between number reading and number identification because of a systematic misunderstanding of the items which causes correlated measurement errors. Children who cannot read the Arabic numerals will struggle with the number identification task as well.

Figure 5.3 shows the prediction model without the autoregressor confirming the results found in predicting Time 2 that the latent variable transcoding, children's understanding and ability to manipulate the Arabic numeral system, was the only unique predictor with an acceptable model-fit to the data, $\chi 2$ (84) = 100.915, p = .101, *RMSEA* = .038 (90% CI = .000 - .062), *CFI* = .963, *SRMR* = .063. Transcoding explained 51% of the variance.

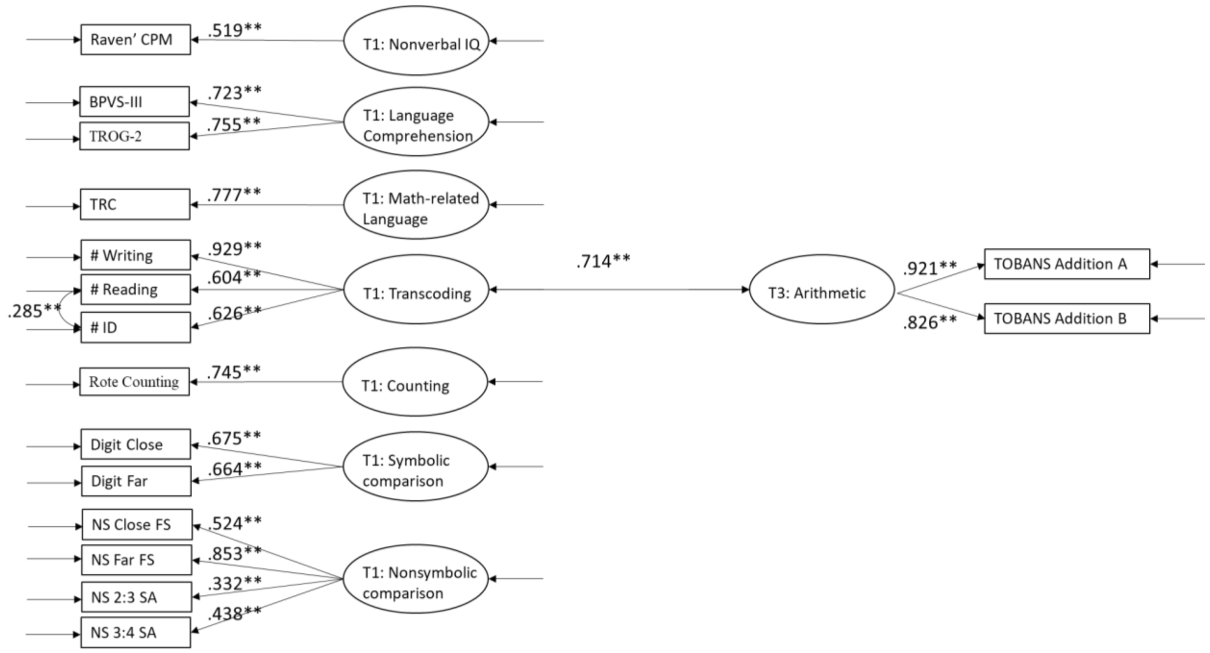


Figure 5.3. Prediction of arithmetic at Time 3 by Time 1 base model without autoregressor. * p < .05. ** p < .01.

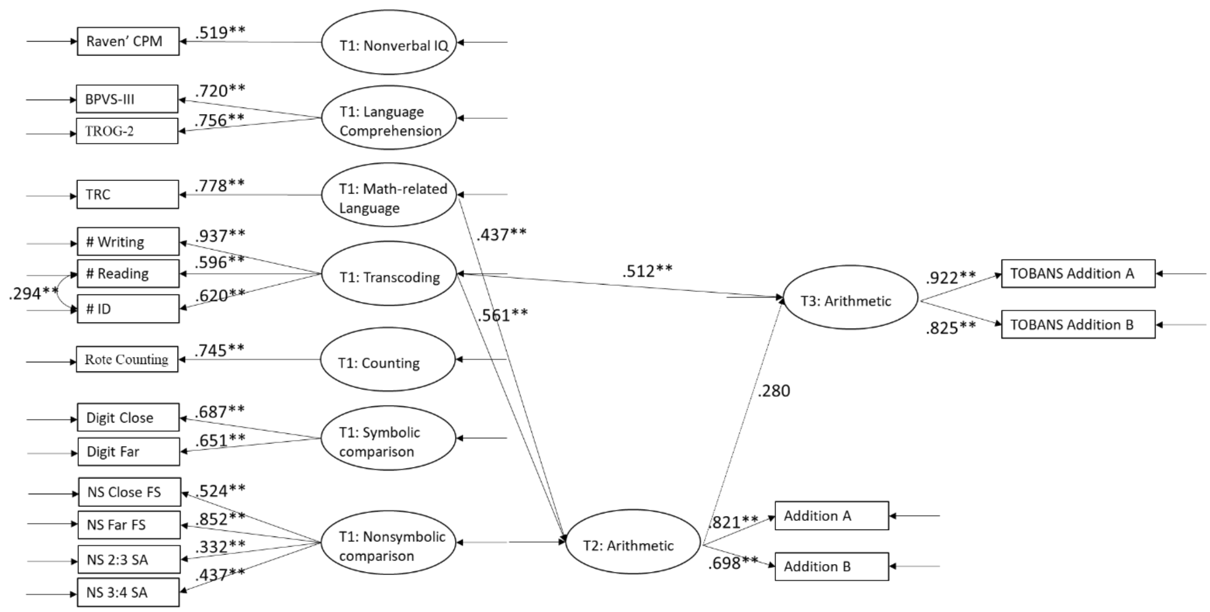


Figure 5.4. Prediction of arithmetic at Time 3 by Time 1 base model with autoregressor * p < .05. ** p < .01.

The model with autoregressor is displayed in Figure 5.4. The model provided an excellent fit to the data, $\chi 2$ (112) = 130.483, *p* = .112, *RMSEA* = .034 (90% CI = .000 - .056), *CFI* = .967, *SRMR* = .067. Similar to the first model, transcoding was the only unique predictor of arithmetic at Time 3 and the autoregressor Time 2 was not a significant predictors (the model explained 53.9% of the variance).

5.2.4 Predicting Time 4.

Similar to Time 3, two models (without autoregressor and with autoregressor Time 2, shown in Figure 5.5 and 5.6) were conducted to investigate the longitudinal prediction of the Time 1 base model (independent variables) onto arithmetic skills at Time 4 (independent variable comprised latent factor of TOBANS subtasks assessed in spring term of Year One, 20 months after Time 1). The path model without the autoregressor provided an adequate fit to the data, $\chi 2$ (100) = 117.411, *p* = .113, *RMSEA* = .035 (90% CI = .000 - .059), *CFI* = .969, *SRMR* = .065. The findings support previous results that children's ability to translate between verbal number codes and Arabic numerals (transcoding) was the only unique predictor of early arithmetic, explaining 45.9% of the variance. The same pattern can be observed when using the autoregressor arithmetic Time 2 (model-fit to data was acceptable, $\chi 2$ (130) = 146.419, *p* = .154, *RMSEA* = .030 (90% CI = .000 - .052, *CFI* = .975, *SRMR* = .068, 47% of variance explained).

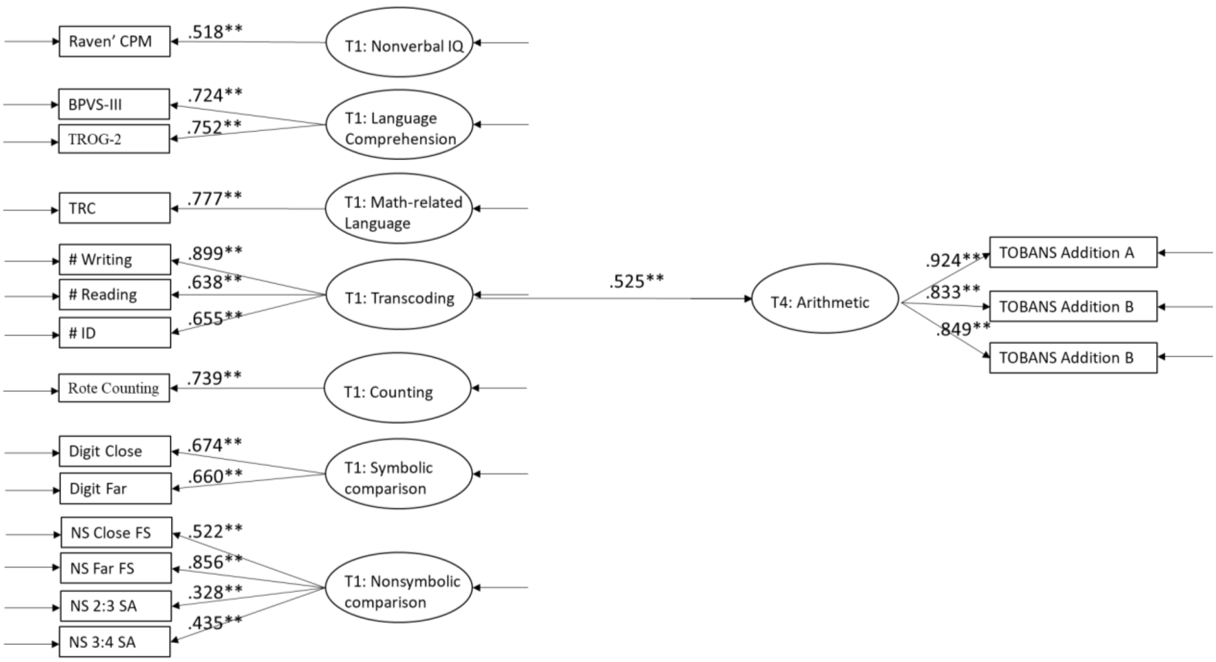


Figure 5.5. Prediction of arithmetic at Time 4 by Time 1 base model without autoregressor.
* p < .05. ** p < .01.

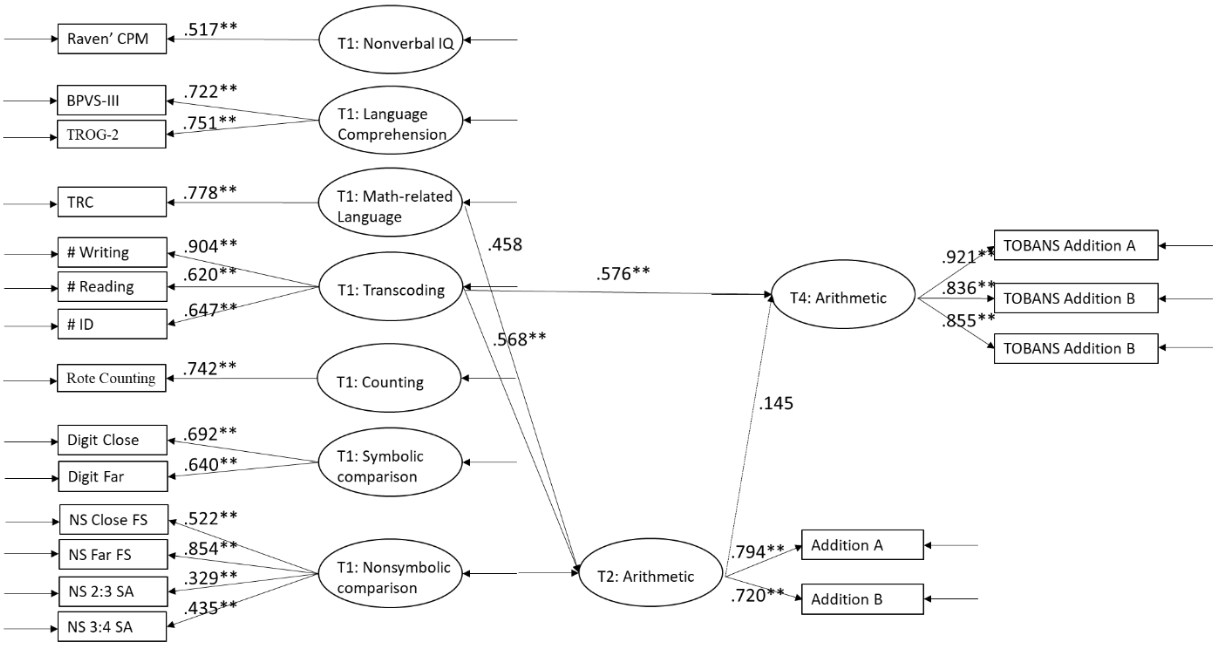


Figure 5.6. Prediction of arithmetic at Time 4 by Time 1 base model with autoregressor * p < .05. ** p < .01.

5.2.5 Predicting Time 5.

The last set of longitudinal prediction models investigated which Time 1 latent, independent variables play an important role in children's performance of arithmetic 25 months later (latent factor consisting of TOBANS subtasks and WIAT-II in summer term of Year One). The nonsymbolic magnitude comparison subtask *surface-area matched* ratio 2:3 was excluded because it loaded poorly on the hypothesised latent variable nonsymbolic magnitude comparison (p = .02).

The SEM path models showed that transcoding uniquely predicted arithmetic corroborating the previous finding that children's understanding of the Arabic numeral system as well as their ability to translate between numerals and verbal codes may be central to the development of early arithmetic skills (48.3% of variance explained). Figure 5.7 depicts the model without the autoregressor (model-fit: $\chi 2$ (100) = 115.579, p = .137, *RMSEA* = .033 (90% CI = .000 - .057), *CFI* = .974, *SRMR* = .070). The model with Time 2 autoregressor provided an acceptable fit to the data, $\chi 2$ (130) = 151.006, p = .100, *RMSEA* = .033 (90% CI = .000 - .054), *CFI* = .969, *SRMR* = .070, 43.2% of variance explained (Figure 5.8). In accordance with previous longitudinal path models, the autoregressor was not a unique significant predictor of arithmetic assessed 25 months later.

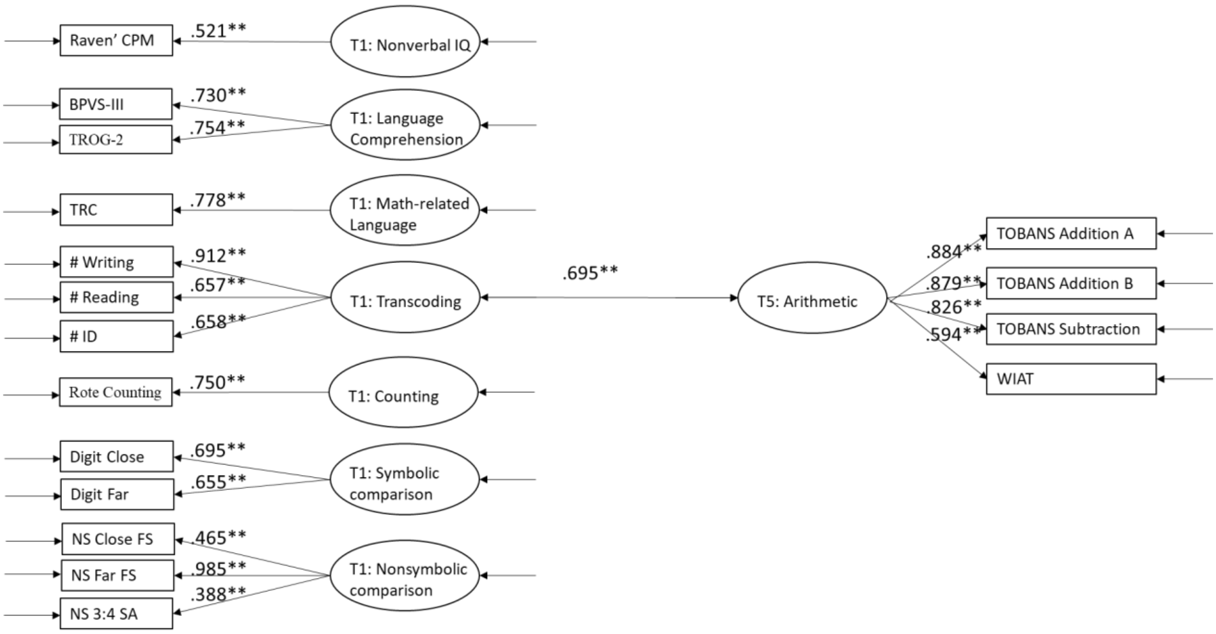


Figure 5.7. Prediction of arithmetic at Time 5 by Time 1 base model without autoregressor. * p < .05. ** p < .01.

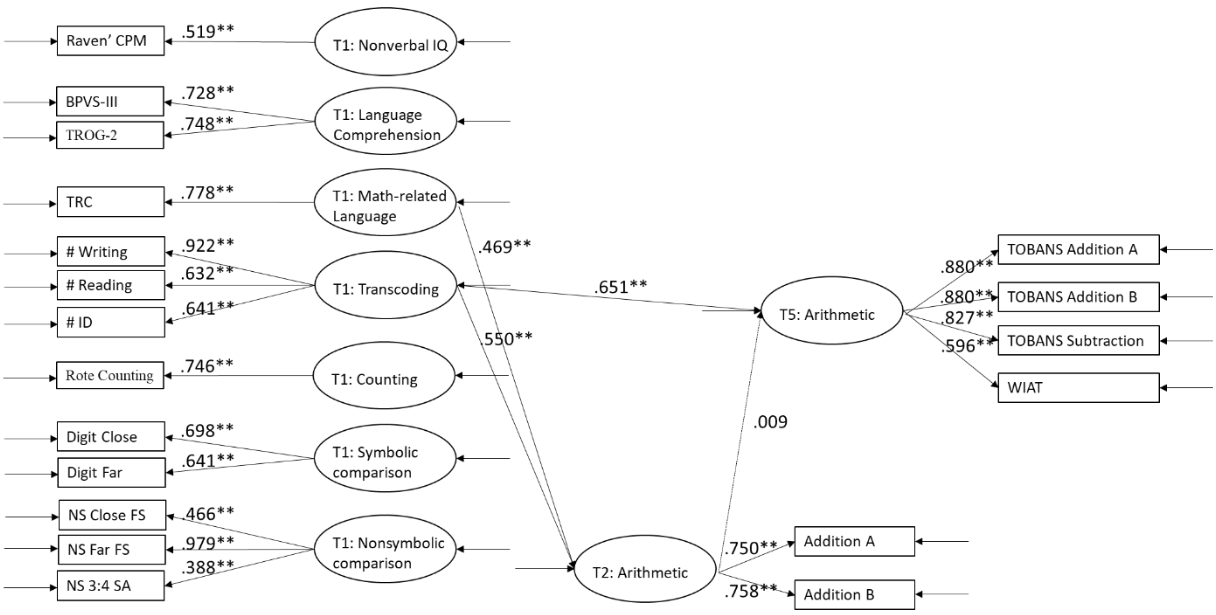


Figure 5.8. Prediction of arithmetic at Time 5 by Time 1 base model with autoregressor * p < .05. ** p < .01.

Similar to the previous chapters, the same Time 5 model with Time 1 arithmetic as autoregressor (Figure 5.9) was conducted only using the data for *number wizards* (high achievers on the number reading task). However, the findings from chapter three suggest that the magnitude comparison tasks should form one factor rather than two distinct factors. Also, *surface-area matched* stimuli with symbolic *far* items were removed due to the fact that they did not significantly load onto the hypothesised general magnitude comparison factor. The number reading task was also removed since this analysis only investigates children who have achieved the maximum score. The model provided an acceptable fit to the data, $\chi 2$ (104) = 113.377, *p* = .249, *RMSEA* = .045 (90% CI = .000 - .092), *CFI* = .962, *SRMR* = .084, with 53.0% of variance explained. Confirming previous results, transcoding and math-related language (both Time 1) comprehension were significant predictors of arithmetic at Time 1 and only transcoding at Time 1 was a significant longitudinal predictor of children's arithmetic scores assessed at Time 5. The prediction from the autoregressor arithmetic at Time 1 was not significant.

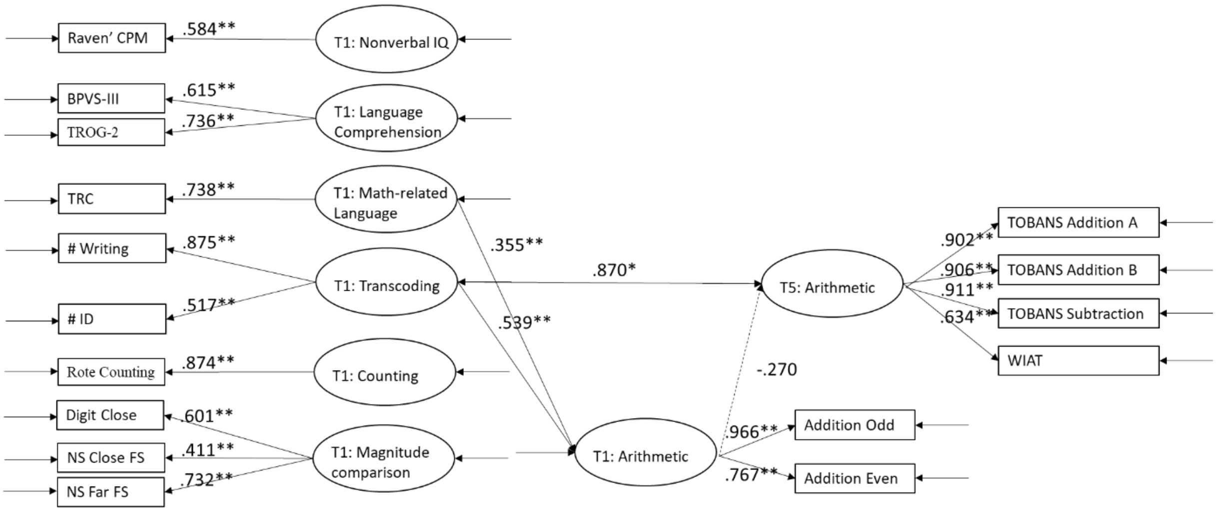


Figure 5.9. Prediction of children's arithmetic performance at Time 5 by Time 1 base model with autoregressor Time 1. The model was run using only the data from high achievers on number reading task at Time 1. * p < .05. ** p < .01.

Correlations. The correlations between the latent constructs are shown in Table 5.2 to 5.5 (for the path models with Time 2 as the autoregressor). The correlation matrices revealed the same correlation pattern across time thus only crucial associations that are stable over time will be discussed.

It seems that the latent outcome variable arithmetic correlated with all other variables, except nonverbal intelligence. Moreover, math-related language comprehension was poorly related with arithmetic and the highest correlation was transcoding which confirmed the findings of prediction of the SEM path models. Furthermore, the variable transcoding correlated highly with counting as well as symbolic comparison whereas the relation to nonsymbolic comparison was fairly weak. It is worth mentioning that counting was associated with transcoding and arithmetic. It seems that numeracy tasks are highly related sharing similar cognitive processes. Interestingly, correlations with nonverbal intelligence were not significant but for nonsymbolic comparison.

5.3 Conclusion.

This chapter explored the longitudinal prediction of early arithmetic in typically developing children over a 25-months period. The main focus was to what extent the ANS and language comprehension specific to mathematical abilities constrain the development of early arithmetic (Libertus et al., 2011; Libertus, Feigenson and Halberda, 2013; Piazza and Dehaene, 2004) and furthermore, to what extent transcoding skills impact early arithmetic. In regards of the results from concurrent prediction of arithmetic (Chapter 4), the findings confirmed that transcoding, children's understanding of the Arabic numeral system, was the only stable longitudinal precursor of early arithmetic skills.

To sum up, transcoding and math-related language comprehension at Time 1 were the only unique longitudinal predictors of children's performance on arithmetic tasks after nine months. It seems that children's early transcoding performance was a slightly stronger predictor of variance in arithmetic tasks than their math-related language comprehension. The autoregressor arithmetic assessed at Time 1 was not uniquely predicting arithmetic. The Time 1 arithmetic measure varied in method and administration from arithmetic at Times 2, 3, 4 and 5. It was not constrained for time and showed ceiling effects.

The results regarding math-related language comprehension support the notion that language impacts early arithmetic. However, most studies have assessed language skills more generally, rather than focusing on language specific to mathematics. Indeed, this thesis may be the first to identify prediction of arithmetic from math-specific language. The findings indicate in particular that understanding of *more*, may be an important foundation in the development of arithmetic.

Table 5.2

Correlations between Time 1 baseline model and arithmetic at Time 2 (n = 142)

	1	2	3	4	5	6	7	8	9
1. Nonverbal Intelligence	---	.303	.173	.380	.194	.321	.596**	.642**	.387*
2. Language		----	.601**	.514**	.283	.603**	.463**	.479**	.573**
3. Math-related Language			---	.323*	.219	.517**	.367**	.448**	.627**
4. Transcoding				----	.672**	.653**	.415**	.625**	.728**
5. Counting					----	.392*	.326*	.393**	.484**
6. Symbolic Comparison						----	.551**	.518**	.615**
7. Nonsymbolic Comparison							----	.518**	.448**
8. Arithmetic Time 1								----	.685**
9. Arithmetic Time 2									----

Notes. Pearson product-moment correlation coefficient. * *p* < .05. ** *p* < .01

Table 5.3

Correlations between Time 1 baseline model with Time 2 autoregressor and arithmetic at Time 3 (n = 143)

	1	2	3	4	5	6	7	8	9
1. Nonverbal Intelligence	---	.238	.246	.397*	.179	.208	.581**	.330	.296*
2. Language		----	.581**	.496**	.271	.594**	.456**	.532**	.403**
3. Math-related Language			---	.302*	.204	.508**	.375**	.607**	.325**
4. Transcoding				----	.666**	.612**	.405**	.693**	.706**
5. Counting					----	.375*	.312*	.463**	.470**
6. Symbolic Comparison						----	.541**	.565**	.471**
7. Nonsymbolic Comparison							----	.391**	.316**
8. Arithmetic Time 2								----	.635**
9. Arithmetic Time 3									----

Notes. Pearson product-moment correlation coefficient. * $p < .05$. ** $p < .0$

Table 5.4

Correlations between Time 1 baseline model with Time 2 autoregressor and arithmetic at Time 4 (n = 143)

	1	2	3	4	5	6	7	8	9
1. Nonverbal Intelligence	---	.231	.241	.415*	.169	.195	.578**	.346*	.289*
2. Language		----	.581**	.479**	.264	.519**	.451**	.538**	.354**
3. Math-related Language			---	.304*	.204	.504**	.372**	.630**	.266**
4. Transcoding				----	.672**	.622**	.397**	.707**	.678**
5. Counting					----	.359*	.304*	.474**	.456**
6. Symbolic Comparison						----	.535**	.584**	.443**
7. Nonsymbolic Comparison							----	.396**	.286**
8. Arithmetic Time 2								----	.552**
9. Arithmetic Time 4									----

Notes. Pearson product-moment correlation coefficient. * $p < .05$. ** $p <$.

Table 5.5

Correlations between Time 1 baseline model with autoregressor Time 2 and arithmetic at Time 5 (n = 148)

	1	2	3	4	5	6	7	8	9
1. Nonverbal Intelligence	---	.236	.240	.410*	.182	.203	.520**	.338	.270*
2. Language		----	.589**	.496**	.276	.598**	.403**	.549**	.328**
3. Math-related Language			---	.306*	.206	.506**	.348**	.637**	.205*
4. Transcoding				----	.671**	.630**	.358**	.693**	.657**
5. Counting					----	.370*	.267	.465**	.441**
6. Symbolic Comparison						----	.496**	.583**	.415**
7. Nonsymbolic Comparison							----	.360**	.236**
8. Arithmetic Time 2								----	.460**
9. Arithmetic Time 5									----

*Notes. Pearson product-moment correlation coefficient. * p < .05. ** p < .0*

A stable pattern emerged that only transcoding at Time 1 was uniquely predicting arithmetic skills after 16-months, 20-months and even 25-months and children's math-related language comprehension may not play an important role in children's performance on arithmetic tasks anymore when investigating the prediction of arithmetic over a longer period. The autoregressor arithmetic assessed at Time 2 were not unique predictors of arithmetic. The findings suggest that children's ability to translate between Arabic numerals and verbal codes substantially impacts the development of early arithmetic skills. The lack of significant prediction from the autoregressor suggests a change in the nature of arithmetic performance at different time points, possibly accounting for the limited timeframe of prediction from math specific language comprehension.

Although the concurrent model showed that nonverbal ability predicted arithmetic skills at Time 1 (Chapter 4), this was not confirmed longitudinally. Results from the concurrent prediction of arithmetic examining only the performance of number wizards (high achievers on the number reading task) found that children's nonverbal intelligence was not a significant predictor of arithmetic. However, nonverbal ability was a unique predictor when analysing the whole sample. These findings suggest that pre-schooler's nonverbal intelligence may not be as important once children's number recognition (number reading) skills are sufficiently developed. This implies that the link between general intelligence and mathematical skills (Cowan et al., 2005; Noël, 2009) may be mediated by early number recognition skills.

Moreover, children's counting scores were a concurrent predictor of early arithmetic at Times 2 and 3 (Chapter 4) confirming a number of previous research findings (Butterworth, 2005; Desoete and Grégoire, 2006; Nunes and Bryant, 1996; Gelman and Gallistel, 1978; Ansari et al., 2003; Cowan et al, 2005; Donlan et al. (2007)). Most studies assessed concurrent prediction rather than longitudinal, though Desoete and Grégoire (2006) reported in their longitudinal study that children performing poorly in arithmetic in grade 1 already struggled in pre-school with number sequence knowledge. The current results reveal that though concurrently predicting arithmetic, children's performance on counting tasks was not a significant longitudinal predictor of arithmetic as suggested in some models of arithmetic development (Aunola et al., 2004; Zhang et al., 2014 and LeFevre et al., 2010)).

Nonetheless, high longitudinal correlations have been reported between counting and arithmetic (Zhang et al., 2014).

Aunola et al. (2004) found that counting was the strongest longitudinal predictor of five-to six-year-olds math achievement. They measured counting in a more complex way than in the current study, including rote counting, counting forwards and backwards from given number and counting in steps. Of particular interest is the fact that Aunola et al. (2004) included a number identification task (similar to measure used in the current study) as part of the outcome measure of maths. It could be that the administration of a broader, more complex counting task as well as the difference in designing the SEM path model (number identification was part of the outcome math measure in Aunola et al. (2004) compared to being a predictor in the current path models) may be reason for the contrasting results.

Also, LeFevre et al. (2010) showed that linguistic and spatial skills form distinct pathways in the development of arithmetic. Linguistic skills contributed to symbolic number system (number naming) and spatial attention skills were related to various math outcome measures (number naming and magnitude comparison). The two links between arithmetic and linguistic skills as well as spatial attention may be mediated by counting sequence knowledge. Zhang et al. (2014) reported in their longitudinal study that children's pre-school letter knowledge and spatial visualisation were predicting first and third grade arithmetic performance. These associations were mediated by counting sequence knowledge assessed in first grade. Further research is necessary to determine the impact of counting on arithmetic development and whether the link between language and math achievement may be mediated by knowledge of the spoken number sequence.

Likewise, measures of the ANS in the cross-sectional analysis (Chapter 4) were significantly predicting children's arithmetic scores at Time 3, 4 and 5. Concerning the longitudinal role of ANS, the findings extend those by Göbel et al. (2014), indicating children's accuracy in magnitude comparison tasks at four years, though strongly correlating with later arithmetic skills, is not a unique predictor of arithmetic skill assessed 25 months later. It was noted that, contrary to Göbel et al. (2014), the symbolic and nonsymbolic magnitude comparison tasks comprised of two independent latent factors. This questions previous findings concerning the involvement of the ANS in early arithmetic development (e.g., Piazza, 2010). Further studies are necessary to clarify the longitudinal role of the ANS in the development of early arithmetic and whether the results hold stable for longer follow-up.

It is important to take account of the finding that neither of the autoregressors (Time 1 and Time 2 arithmetic performance) significantly predicted arithmetic at later stages. This could be due to the fact that the testing procedure for early arithmetic was changed over the course of the study to adjust for children's growing learning experience in arithmetic. In particular, there were ceiling effects in arithmetic scores assessed at Time 1, which children could complete in their own time. At Time 2, time to solve arithmetic was limited to three minutes, but the time given may have been too long to achieve high sensitivity at this age. It seems that the TOBANS was a sensitive measure of arithmetic and it may be interesting to see in future studies if children as young as four years can successfully perform on the TOBANS.

Also, the Time 1 counting measure (highest number produced in correct order) produced high variability in scores. A more complex measure such as the composite of different counting tasks used by Aunola et al. (2004) may prove to be useful in further examining the relationship between ANS, counting, transcoding and arithmetic.

A further constraint of the study was the relatively small sample size which makes it impractical to investigate more complex SEM path models including additional covariates of math achievement such as children's early memory or spatial skills. Further large scale studies are needed to clarify longitudinal prediction using more sensitive and comprehensive measures to enable more detailed conclusions about prediction of early arithmetic, and to ascertain whether the findings hold stable for longer-term follow-up.

Chapter 6. Relations between Inhibitory Control, Approximate Number System and Early Arithmetic

A number of studies provide evidence for the link between the ANS and math achievement (Halberda et al., 2008, Libertus et al., 2013 and Piazza et al., 2010). Some researchers argue that it is possible that this link may not be driven by numerical processing but inhibition skills, based on the findings that children's inhibition skills are reported to also strongly relate to math learning (Gilmore et al., 2013, Gilmore et al., 2014 and Fuhs and McNeill, 2013). McCelland, Cameron, Connor, Farris, Jewkes and Morrison (2007) reported that four-year-olds' behavioural regulation scores were significantly and positively correlated with their literacy, vocabulary and math skills after five months.

In a recent study, Gilmore et al. (2013) showed that children performing nonsymbolic magnitude comparison tasks were less accurate on incongruent trials (where dot size and envelope area are negatively correlated with number of dots) than congruent trials (where dot size and envelope area are positively correlated with number of dots) suggesting that to solve incongruent problems children had to draw on the additional processing step of inhibitory control. Furthermore, the results showed that incongruent and not congruent items correlated significantly with maths achievement. They further argued that the correlation found between maths and dot comparison was driven by incongruent trials and hence children's inhibition skills. In a second experiment they found that children's performance on ANS tasks did not significantly predict math achievement once inhibition skills had been accounted for supporting the hypothesis that the relationship between ANS and arithmetic is not driven by the nature of underlying numerical representations but by inhibitory control demands of some dot comparison trials. However, the age range of the study was large (5 to 12 years in first experiment and 8 to11 years in second experiment), and children's age was not taken into account.

In addition, Fuhs and McNeill (2013) showed that nonsymbolic magnitude comparison was predicting children's mathematical skills, but nonsymbolic comparison was not a predictor once inhibition was taken into account. All of these studies only investigated inhibition tasks and ANS neglecting other important covariates of math learning thus risking false attribution of causation. Gilmore et al. (2013) argues that their study has the benefit of including a naming task as well as an inhibition measure, but this is still a limited model. Fuhs and McNeil (2013) have a low income sample but it is substantial. The age range of four to six may seem broad, but they controlled for age and the analysis was more comprehensive, though still failing to take account of children's knowledge of symbolic versus nonsymbolic magnitude comparison.

To answer the research question concerning the relationship between inhibition, ANS and early arithmetic, the current chapter assesses to what extent children's performance on early arithmetic tasks can be predicted by magnitude comparison and to what extent this relationship is driven by children's inhibitory control. The same behavioural regulation task as used by McClelland et al. (2007) was administered. The analyses will only investigate performance on ANS and inhibition, although findings from Chapters 4 and 5 suggest that neither of the two measures may be crucial in math learning when a number of covariates such as transcoding are taken into account.

First, children's performance on congruent and incongruent dot comparison items will be examined as well as their relationship to early arithmetic, replicating the first experiment of Gilmore et al. (2013). The second part of the chapter will replicate the second experiment of Gilmore et al. (2013), analysing the link between inhibition, ANS and arithmetic.

6.1 Congruent versus Incongruent ANS Trials.

The experiment focuses on children's performance on congruent (smaller array of dots covers a smaller area compared to larger array of dots with a larger surface; fixed size condition from Chapter 3) and incongruent (smaller array of dots covers larger area compared to larger array of dots with small surface; surface-area matched condition from Chapter 3) magnitude comparison tasks. The relationships between early arithmetic and congruent or incongruent items were examined to establish which of the two may be important in the development of arithmetic.

6.1.1 Method.

6.1.1.1 Participants.

The same participants were used as described in Chapter 2 (p. 42)

6.1.1.2 Materials.

6.1.1.2.1 Measures taken at Times 1 and 2.

Magnitude Comparison. Various nonsymbolic comparison tasks were created for the study, based on those used in Göbel et al. (2014). Each comparison pair was presented on a single page (see Figure 6.1). Children were given one point for every correct comparison with a maximum score of 16 for each version and 160 overall (for more details see Chapter 3, pp.50). The experimenter made sure that children did not count the dots.

Congruent trials were the trials called *fixed size*. Because the size of dots was fixed, the larger array of dots was also the array with the larger area printed in black. Contrary, the incongruent items were the *surface-area matched* items where the area printed in black was the same, meaning the smaller array had bigger dots compared to the larger array.

Arithmetic Skills. The children's basic arithmetic skills at Time 1 were assessed using simple addition problems. The test comprised ten simple additions with sums less than ten (1 + 3; 2 + 1; 2 + 2; 1 + 4; 3 + 1; 1 + 5; 2 + 3; 1 + 6; 3 + 3; 4 + 4). All arithmetic problems were presented in Arabic notation (MS Office 2013, Comic Sans MS, size 260) and, simultaneously, in spoken form most familiar to the child. Problems were arranged so that additions with same sums or similar summands were never adjacent. Children were encouraged to use wooden sticks provided or their fingers if needed. Before two practice problems (1 + 1, 1 + 2) were administered, the preferred method of referring to additions ("add" or "plus") was determined by asking the teachers. Testing was only terminated early if a child showed signs of confusion or lack of concentration. The maximum score was ten.

Due to the ceiling effect performance at Time 1, the following adjustments were made to the basic calculation task for Time 2: Two parallel forms of the tasks have been created which comprised of ten simple additions with sums less than ten (both forms were equal in difficulty level; Form A: 1 + 3; 2 + 1; 1 + 5; 2 + 3; 4 + 5; 7 + 2; 3 + 5; 4 + 2; 5 + 2 and 2 + 6; Form B: 1 + 4; 3 + 1; 2 + 5; 4 + 2; 1 + 6; 3 + 6; 2 + 7; 6 + 2; 4 + 3 and 3 + 5). To raise the sensitivity of the task even further, children had only three minutes to solve as many problems as possible. The two forms were given in two separate testing sessions. To avoid training effects, the order of the forms was counterbalanced. The total number of correctly solved problems was recorded.

6.1.1.2.2 Measures taken at Times 3, 4 and 5.

Magnitude Comparison. A recent study by (Göbel et al., 2014) showed that children in Year One can successfully perform magnitude comparison tasks in a group setting. Thus the magnitude comparison task used in this study was redesigned as a group test using the same stimuli pairs created at Times 1 and 2. Nonsymbolic comparisons were presented in pairs of two adjacent 2.1 cm x 2.1 cm boxes. Children were asked to tick the bigger number or box with more dots (for more details see Chapter 3, pp.50).

Arithmetic Skills. *Fluency.* Children's speeded arithmetic skills (fluency) at Time 3 was assessed using the 'addition' and 'addition with carry' subtests of the Test of Basic Arithmetic and Numeracy Skills (TOBANS; Brigstocke et al., 2016). Children were asked to complete as many arithmetic problems as possible in one minute. In the 'addition' subtask, children were presented with simple addition problems with sums less than ten and in the 'addition with carry' subtask the sums were bigger than ten but smaller than twenty. One point was awarded even if the numeral was written backwards (maximum score $_{\text{addition}}$ = 90; maximum score $_{\text{addition with carry}}$ = 30). This task was administered as a group task.

At Times 4 and 5, children were also presented with the 'subtraction' subtask of the TOBANS in addition to the 'addition' and 'addition with carry' subtasks. Similar to 'addition', children were asked to solve as many of the 90 subtraction problems as possible in one minute.

Accuracy. Children's basic arithmetic accuracy was assessed using the Numerical Operations subtest of the second edition of the Wechsler Individual Achievement Test (WIAT-II; Wechsler, 2005) at Time 5. The first six items (identifying and writing Arabic numerals) were excluded because we were only interested in a more conventional measure of arithmetic. The test was executed according to the manual and children were allowed to complete the task in their own time (maximum score = 25).

6.1.1.3 Procedure.

The ANS and arithmetic tasks were part of a comprehensive test battery. Testing was carried out five times over a 25-month period from the summer term of nursery through to the summer term of Year One. Wherever possible, each child was seen by the same experimenter two to four days in a row. The main researcher was assisted by several research assistants from undergraduate psychology classes. They were trained on how to administer the test battery and were given instructions on how to work with young children. Children were tested in a separate room or another quiet place in the school.

Arithmetic was assessed individually at Times 1 and 2 and in groups at Times 3, 4 and 5. Children could solve the addition problems at Time 1 on their own time, whereas the Times 2, 3, 4 and 5 arithmetic tasks were time limited. Similarly, the magnitude comparison task at Times 1 and 2 were individually assessed and children were allowed to finish the tasks on their own time. The magnitude comparison task was re-designed as a group task at Time 3 and was hence forward assessed in a group setting with a ratio of experimenters to children of 1:3 (for order of fixed size and surface-area matched subtasks see Appendix 19).

Preliminary to the testing, the experimenters attended at least one day in each class so that the children got to know them and felt more comfortable around them. Moreover, the experimenters told each child that they would play games and asked questions such as “How are you?” or “How old are you?”.

All unstandardized tests included practice items. Concerning feedback, children received only concrete feedback on their performance for practice items and general praise and encouragement throughout the tests.

Table 6.1

Mean and standard deviations of predictor and criterion measures from all testing sessions

		Time 1	Time 2	Time 3	Time 4	Time 5
		M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Behavioural regulation	HTSK			A: 14.71 (4.41) [9] B: 10.30 (5.90) ([3]*	A: 17.25 (3.00) [16] B: 15.47 (4.66) [16]*	
Magnitude Comparison	Congruent (FS)	33.89 (5.87)	61.15 (7.85)	57.74 (21.54)	75.32 (20.19)	84.92 (25.42)
	Incongruent (SA)	34.34 (4.98)	58.54 (7.31)	53.21 (20.20)	68.83 (18.69)	77.45 (23.11)
		10.11 (3.21) [7]	11.97 (3.07) [20]	10.90 (4.92)	13.04 (4.46)	16.89 (4.57)
	Digit Close	10.73 (3.59) [7]	13.64 (3.33) [60]	14.32 (4.85)	17.20 (5.33)	22.32 (5.82) [2]
	Digit Far	10.26 (2.16) [1]	10.64 (2.20) [1]	9.99 (4.14)	12.12 (4.07)	15.14 (4.33)
	NS FS Close	12.87 (2.57) [17]	14.13 (2.04) [34]	15.49 (6.11)	17.99 (5.31)	21.68 (6.32) [3]
	NS FS Far		13.18 (2.35) [24]	13.90 (5.12)	17.63 (5.98)	21.98 (6.41) [4]
	NS FS 2:3	10.76 (2.72) [4]	11.97 (2.44) [16]	12.82 (5.38)	15.56 (5.23)	19.68 (5.86) [2]
	NS FS 3:4	10.58 (2.29) [3]	11.23 (2.28) [3]	9.86 (4.95)	12.40 (3.94)	14.81 (4.95)
	NS FS 5:6	10.29 (2.28) [1]	10.67 (2.00)	8.99 (3.64)	10.57 (3.77)	12.47 (3.93)

	NS SA Close	13.24 (2.43) [23]	13.49 (2.09) [19]	14.48 (5.93)	17.78 (5.45)	21.96 (6.30) [2]
	NS SA Far	11.42 (2.42) [5]	12.37 (2.28) [8]	12.68 (5.51)	16.14 (5.42)	19.12 (6.45)
	NS SA 2:3	10.81 (2.03) [2]*	11.56 (2.23) [3]	10.92 (5.14)	13.40 (4.90)	17.35 (5.45)
	NS SA 3:4		10.45 (2.11) [1]*	9.60 (4.13)	11.34 (3.73)	13.23 (4.60)
	NS SA 5:6					
Arithmetic	Addition Tasks	6.33; 3.21 (17)*	A: 5.23; 2.51 (3) B: 5.15; 2.55 (8)*			Addition: 6.02; 1.51 (10) Subtraction: 5.04; 2.41 (10)
	TOBANS					
	Addition			6.23; 4.55	8.36; 5.09	12.74; 8.66
	Addition w/			1.75; 2.20	2.56;2.74	5.07; 5.01 (1)
	carry				5.30; 4.12	8.44; 5.10
	Subtraction					
	WIAT					4.00; 2.27

*Notes. M = mean age. SD = standard deviation * individually administered tasks. The number of children scoring the maximum score are shown in square brackets. All scores are presented as raw scores. For the Magnitude Comparison Tasks: NS = nonsymbolic. FS = fixed size trials. SA = surface-area matched trials.*

6.1.2 Results.

The descriptive statistics of the congruent (*fixed size*) items and incongruent (*surface-area matched*) items are shown in Table 6.1. To answer the research question regarding the relationship between congruency and early arithmetic performance, it was decided to focus on inhibition and ANS only, to see whether the findings from Gilmore et al. (2013) can be replicated. More complex analyses of predictors of arithmetic are discussed in Chapters 4 and 5. Descriptive statistics are examined first, followed by simple linear regression run on IBM SPSS Statistics 22 (dependent variable comprised of the composite score of arithmetic subtask raw scores and independent variables were composite scores for incongruent and congruent items from the magnitude comparison task), followed by structural equation modelling using MPlus Version 7. In the SEM path models, the dependent variable was the latent factor arithmetic and the independent variables were the latent factors for congruent and incongruent trials. Congruent trials were the trials called *fixed size*. In contrast, the incongruent items were the *surface-area matched* items.

6.1.2.1 Time 1.

Children answered 53.66% of the incongruent trials correctly (32 children scored below chance level), and 52.95% of congruent (42 children scored below chance level). On average, children did not perform significantly better on incongruent trials (M = 34.34, SD = 4.98) than congruent trials (M = 34.34, SD = 4.98), $t(99) = -.814$, $p = .417$. The correlation between congruent and incongruent conditions was significant, $r = .528$. As shown in scatterplots (Figure 6.1), relationship between congruent items and early arithmetic at Time 1 ($r^2_{congruent} = .145$) was somewhat stronger than the relation between incongruent items and arithmetic ($r^2_{incongruent} = .098$). Congruent trials significantly predicted arithmetic scores, $ß= .30$, $t(97) = 2.81$, $p = .006$, contrary to incongruent trials, $ß = .17$, $t(97) = 1.55$, $p = .124$.

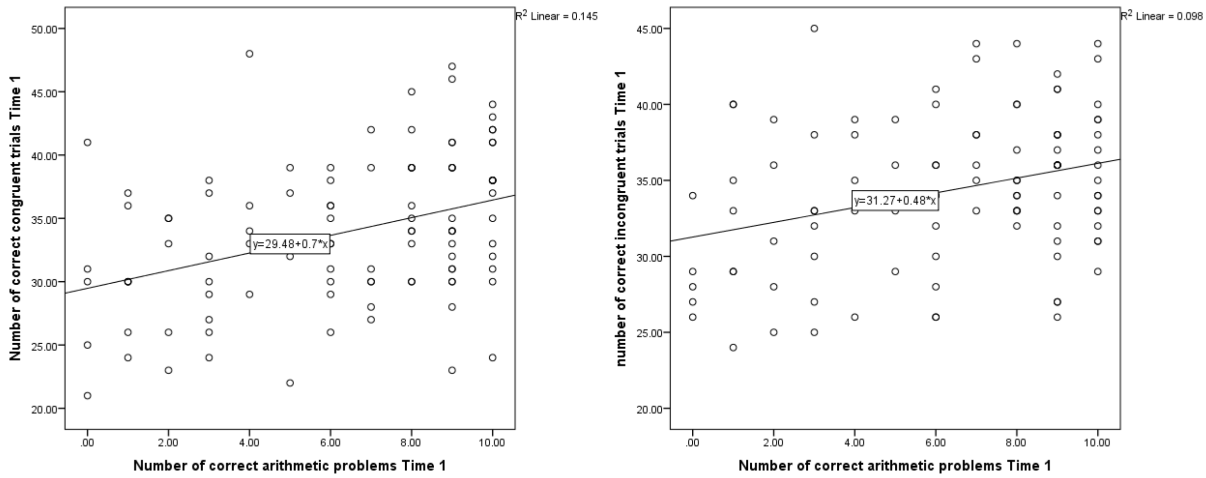


Figure 6.1. Scatterdot plot of congruent trial performance and arithmetic skills (r^2 = .145 left side) compared to incongruent trial performance and arithmetic skills at Time 1 (r^2 = .098, right side).

The second analysis focused on investigating the relationship between congruency of magnitude comparison items and children's performance on early arithmetic tasks using structural equation modelling. The path model (Figure 6.2) shows that neither congruent nor incongruent trials were significant predictors of arithmetic at Time 1. The model fit was excellent, $\chi 2$ (17) = 18.636, *p* = .350, *RMSEA* = .031 (90% CI = .000 - .098), *CFI* = .993, *SRMR* = .041, 25% of variance explained.

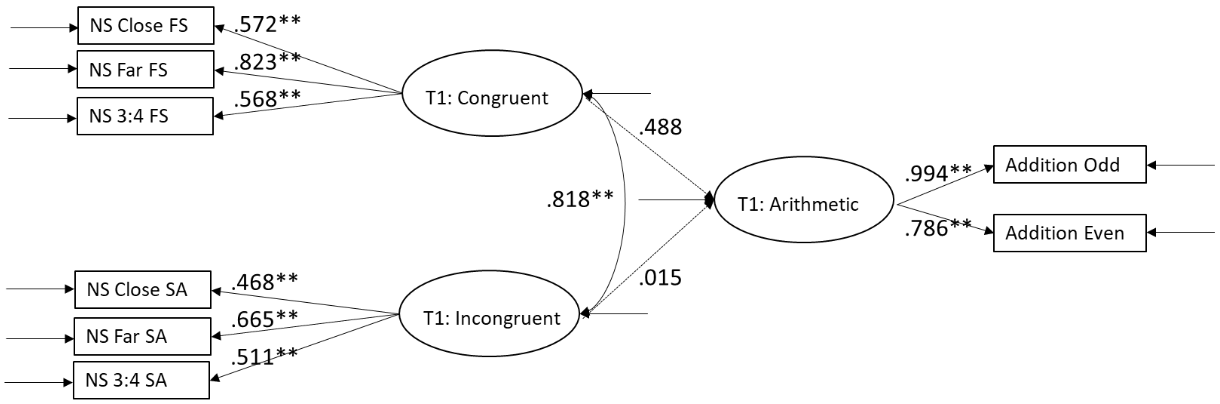


Figure 6.2. Prediction of arithmetic scores by congruent and incongruent latent factors at Time 1.

6.1.2.2 Time 2.

Children's performance at Time 2 was more accurate than Time 1, with 76.44% of correctly answered congruent trials compared to 73.18% of correctly answered incongruent trials. Contrary to Time 1, children performed significantly better on congruent trials (*M* = 61.15, *SD* = 7.85) than incongruent trials (*M* = 58.54, *SD* = 7.31), $t(116) = 4.807$, $p < .001$. The correlation between congruent and incongruent conditions was highly significant, with $r = .703$.

The relationship between congruent items and early arithmetic at Time 2 ($r^2_{congruent} = .189$) was stronger than the relationship between incongruent items and arithmetic at Time 2 ($r^2_{incongruent} = .131$) as indicated in scatterplot graphs (Figure 6.3). Similarly, congruent trials significantly predicted arithmetic scores, $ß = .36$, $t(114) = 3.009$, $p = .003$, in contrast to incongruent trials, $ß = .11$, $t(114) = .950$, $p = .344$.

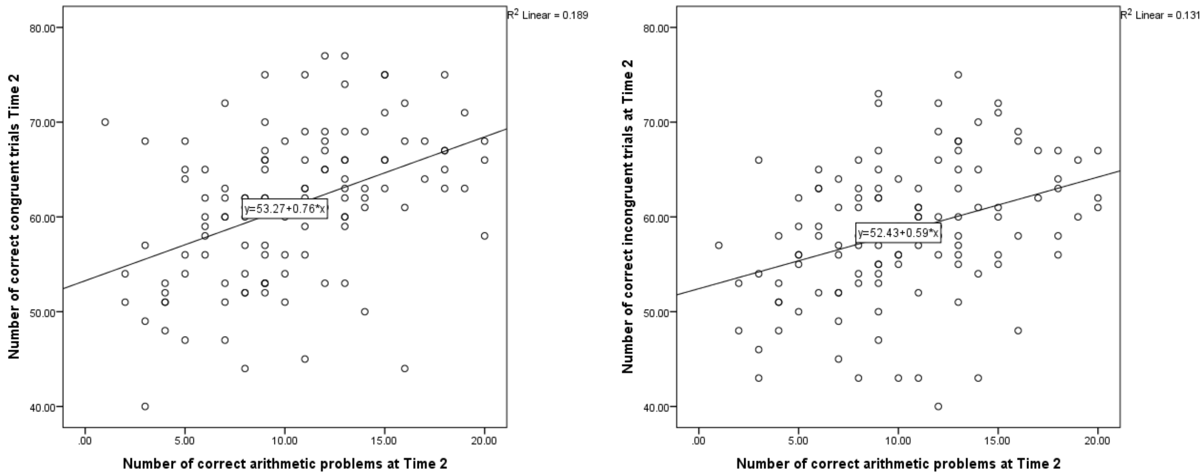


Figure 6.3. Scatterdot plot of congruent trial performance and arithmetic skills ($r^2 = .189$, left side) compared to incongruent trial performance and arithmetic skills at Time 2 ($r^2 = .131$, right side).

The SEM path model examining the relationship between congruency of magnitude comparison items and children's performance on early arithmetic tasks at Time 2 provided an excellent fit to the data, $\chi 2$ (51) = 58.438, $p = .221$, *RMSEA* = .035 (90% CI = .000 - .072), *CFI* = .980, *SRMR* = .050, 47.3% of variance explained (Figure 6.4). Again, neither the congruent nor the incongruent factor predicted children's arithmetic scores, but the congruent factor was the strongest predictor. The model estimation terminated normally, nonetheless, parameters from congruent and incongruent factors to arithmetic factor were both greater than 1, most likely due to the high correlation between congruent and incongruent trials ($r = .977$).

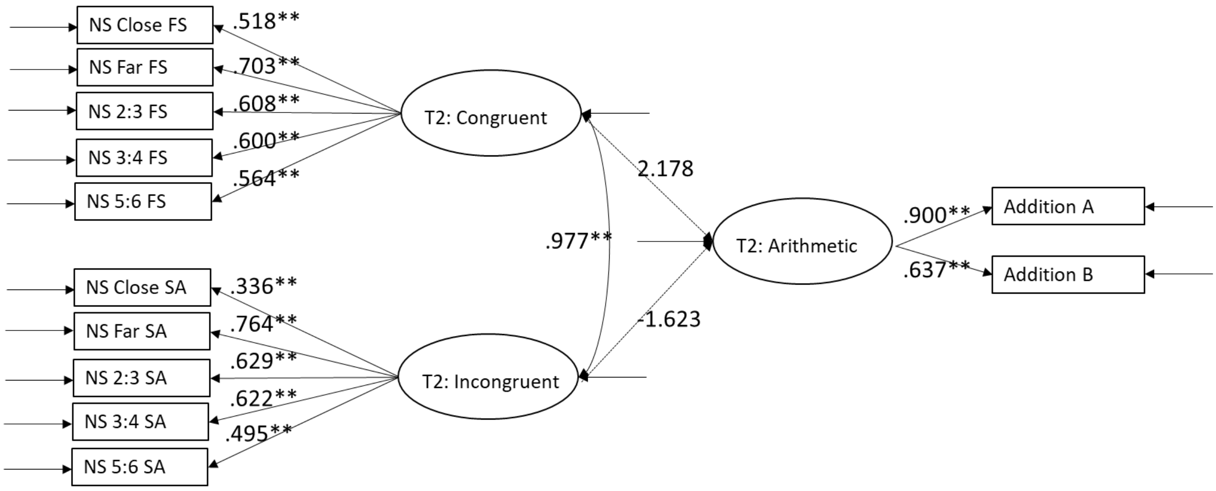


Figure 6.4. Prediction of arithmetic scores by congruent and incongruent latent factors at Time 2.

6.1.2.3 Time 3.

Children's accuracy at Time 3 increased further with 83.30% of accuracy on congruent trials and 76.16% of accuracy on incongruent trials. The performance on congruent trials (M = 57.74, SD = 21.54) was significantly better than the performance on incongruent trials (M = 53.21, SD = 20.20), $t(115) = 6.055$, $p < .001$. Congruent and incongruent conditions correlated highly, $r = .927$. The relationship between congruent items and arithmetic scores at Time 3 substantially improved compared to Time 2 ($r^2_{congruent} = .378$). The relationship between incongruent items and arithmetic at Time 3 ($r^2_{incongruent} = .360$) also substantially improved over time (Figure 6.5). Arithmetic performance at Time 3 was significantly predicted by congruent items, $ß = .419$, $t(113) = 2.127$, $p = .036$, but not incongruent trials, $ß = .211$, $t(113) = 1.07$, $p = .287$.

Running the SEM path model on prediction of arithmetic at Time 3 through MPlus 7 resulted in the warning that the latent variable covariance matrix was not positive defined. Examining the correlation matrix showed that congruent and incongruent magnitude comparison factors were linear dependent, with $r = 1.035$ suggesting that they form one unitary factor and not two distinct factors. Hence, further structural equation modelling was deemed to be invalid for Time 3.

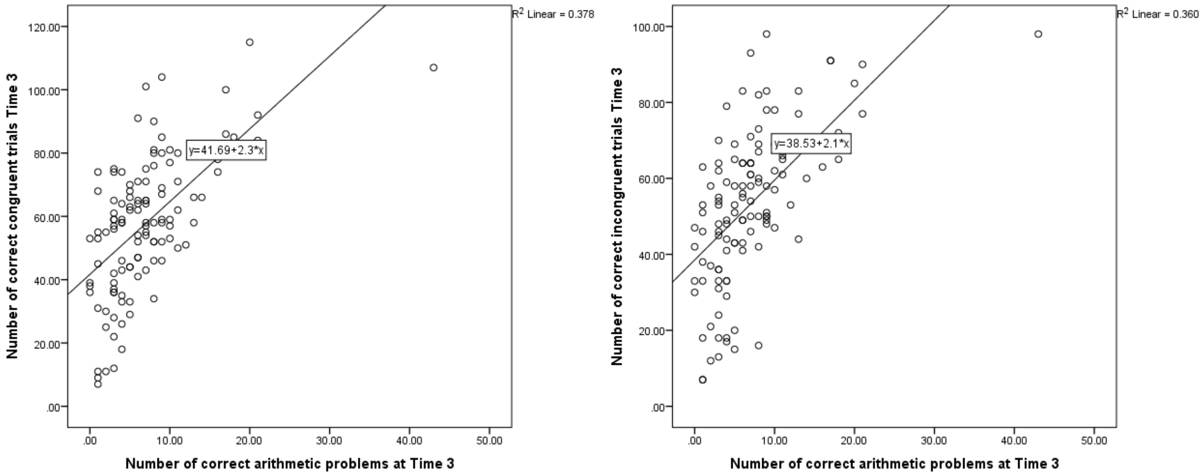


Figure 6.5. Scatterdot plot of congruent trial performance and arithmetic skills (r^2 = .378, left side) compared to incongruent trial performance and arithmetic skills at Time 3 (r^2 = .360, right side).

6.1.2.4 Time 4.

Similar to previous testing points, performance on congruent trials (*M* = 75.32, *SD* = 20.19) was significantly different from the performance on incongruent trials (*M* = 68.83, *SD* = 18.69), $t(110) = 8.663$, $p < .001$ and congruent and incongruent conditions correlated highly, $r = .920$. At Time 4, the relationship between congruent trials and arithmetic ($r^2_{congruent} = .332$) was similar to the relationship between incongruent trials and arithmetic ($r^2_{incongruent} = .309$), and comparable to Time 3. The relationships are shown as scatterplot graphs in Figure 6.6.

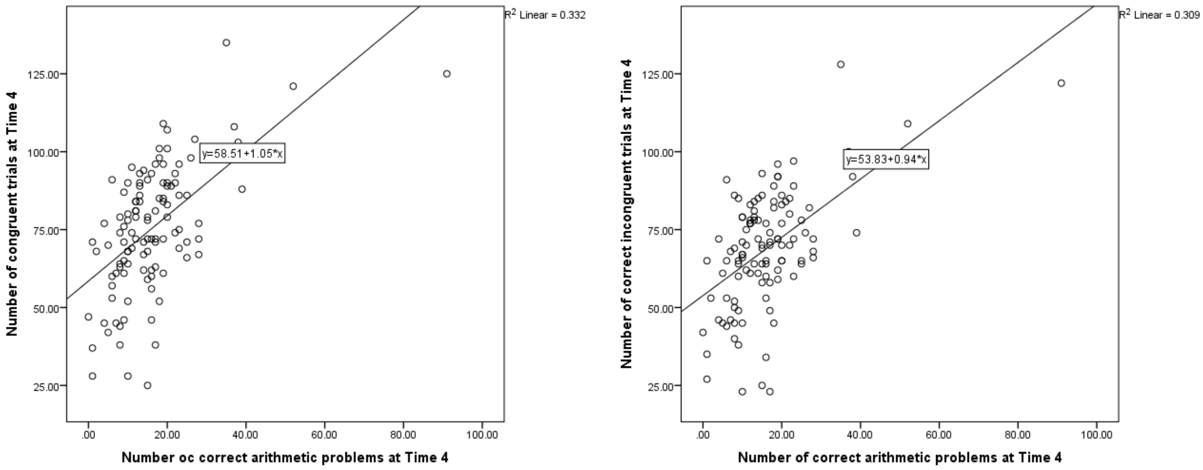


Figure 6.6. Scatterdot plot of congruent trial performance and arithmetic skills (r^2 = .332, left side) compared to incongruent trial performance and arithmetic skills at Time 4 (r^2 = .309, right side).

Congruent trials were significantly predicting children's arithmetic scores at Time 4, *ß* = .425, *t*(108) = 2.117, *p* = .037. Incongruent trials did not predict arithmetic, *ß* = .165, *t*(108) = .822, *p* = .413.

Similar to Time 3, MPlus 7 running SEM path models resulted in a latent variable covariance matrix not positively defined, due to linear dependency between congruent and incongruent trials, *r* = 1.045. Again, further structural equation modelling was deemed to be invalid and was abandoned.

6.1.2.5 Time 5.

Children performed significantly better on congruent trials (*M* = 84.92, *SD* = 25.42) compared to incongruent trials (*M* = 77.45, *SD* = 23.11), *t*(116) = 8.101, *p* < .001, *r* = .920).

At Time 5, congruent trials and arithmetic scores were moderately related ($r^2_{congruent}$ = .244), similar to the relationship between incongruent items and arithmetic performance ($r^2_{incongruent}$ = .230). Scatterplot graphs depicting the relationships are shown in Figure 6.7. Interestingly, neither congruent items (*ß* = .339, *t*(114) = 1.639, *p* = .104) nor incongruent trials (*ß* = .167, *t*(114) = .808, *p* = .421) were significant predictors of children's arithmetic scores at Time 5.

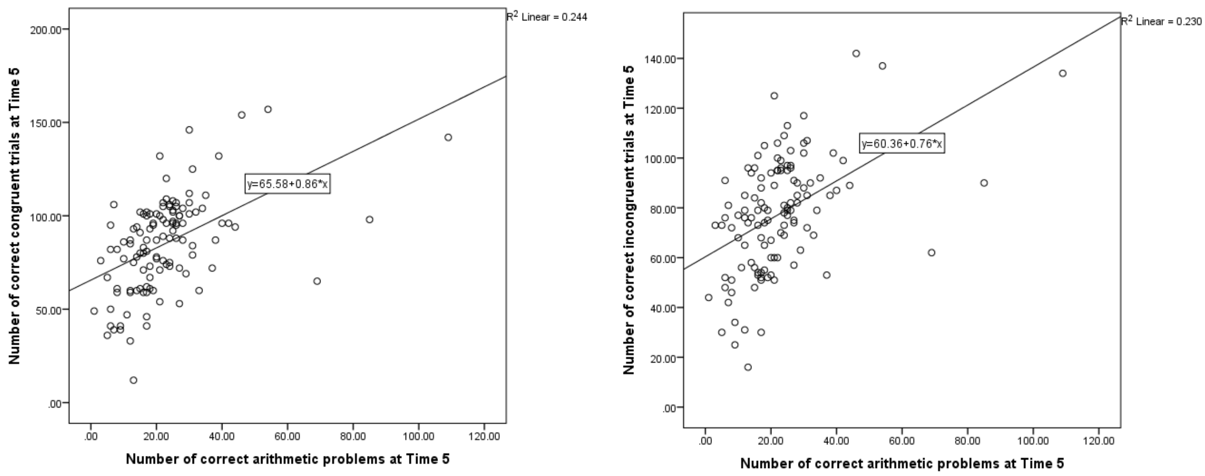


Figure 6.7. Scatterdot plot of congruent trial performance and arithmetic skills (r^2 = .244, left side) compared to incongruent trial performance and arithmetic skills at Time 5 (r^2 = .230, right side).

As seen in the previous two testing points, running the SEM path model on prediction of arithmetic at Time 5 resulted in a latent variable covariance matrix not positively defined. The correlation matrix showed that congruent and incongruent factors were linear dependent, with $r = 1.000$ suggesting that they form one unitary factor and not two distinct factors. Hence, further structural equation modelling was deemed to be invalid for Time 5.

6.1.3 Conclusion.

Overall, children's performance on congruent as well as incongruent items improved over time. They performed significantly better on the congruent condition than the incongruent condition replicating the findings by Gilmore et al. (2013) at all time-points except for Time 1. At Time 1, at least a third of children performed below chance level, though by Time 2 all of the children scored above chance level. It may be that, as Gilmore suggests, solving incongruent items is harder for children due to the added cognitive process of inhibiting the salient, but not useful feature of dot size. However, the results failed to reproduce the finding that performance on incongruent items predicted arithmetic.

Gilmore et al. (2013) argued that relationships between dot comparison and arithmetic are explained by inhibition and that the underlying relationship between inhibition processes and arithmetic is the sole driver for the link between ANS and arithmetic. For Gilmore, this was demonstrated by the finding that incongruent, but not congruent trials, are correlated with arithmetic. The results of this study showed the reverse that congruent (but not incongruent) condition scores predicted children's arithmetic scores in linear regressions. This pattern emerged across all time points except for Time 5 where neither congruent nor incongruent items predicted arithmetic. It seems that children's inhibition skill may not be as crucial in early arithmetic as shown by Gilmore et al. (2013). This was confirmed through SEM path models for at least Times 1 and 2. Again, neither predicted children's arithmetic scores though accuracy on congruent trials showed a stronger relation to early arithmetic. After Time 3, the latent congruent and incongruent factors were highly correlated suggesting the two form one factor and should not be considered to be distinct constructs. Thus it seems that both congruent and incongruent trials access a common representation after five years ten months of age.

There are various reasons why our study failed to replicate Gilmore et al.'s (2013) findings: One major difference between the studies was the age range. This study focused on a narrow age range critical to the period of arithmetic development. Gilmore and colleagues assessed a wider age range but failed to control for age. Furthermore, the congruent and incongruent trials were not the exact same stimuli as used in Gilmore et al. (2013).

6.2 Inhibitory Control, ANS and Arithmetic.

The next experiment assessed the relationship between ANS, inhibition and arithmetic (second experiment of Gilmore et al. (2013)).

6.2.1 Method.

6.2.1.1 Participants.

The same participants were used as described in Chapter 2 (p. 42)

6.2.1.2 Materials.

Children were assessed on the following measures at Times 3 and 4.

Inhibition. To assess children's inhibition skills at Times 3 and 4, we individually administered the Head-to-Toes task (Cameron Ponitz, McClelland, Jewkes, McDonald Connor, Farris and Morrison, 2008; Appendix 16)). The task requires the children to do the opposite of what the experimenter asks them to do. If children were asked to touch their head (or their toes), the correct response would be to touch the toes (head). The experimenter demonstrated the task to the child before four practice items were administered where instructions were repeated up to three times. After the practice items, the test items were executed comprising of one block with ten head-toe items. One point was awarded if the child had to self-correct the answer (child first moves to incorrect response but then stops and response correctly) and two points were given if a child gave the correct response without hesitation or a prior movement to the incorrect response. The second block was administered if the child responded correctly to at least five test trials. This block involved a second set of commands - if the child was asked to touch their shoulders (or knees), they had to touch their knees (or shoulders). Similar to the first block, the experimenter first demonstrated the task before administering the four practise items. After the practice items, ten further test trials were given with commands from the first block mixed with the new commands. The total score possible on each block was 20 points and the maximum overall score was 40 points.

Magnitude Comparison. The magnitude comparison task used in this study was assessed as a group test (see Göbel et al., 2014) using the same stimuli pairs created at Times 1 and 2. Nonsymbolic comparisons were presented in pairs of two adjacent 2.1 cm x 2.1 cm boxes. Children were asked to tick the bigger number or box with more dots (for more details see Chapter 3, pp.50).

Arithmetic Skills. Children's speeded arithmetic skills (fluency) at Time 3 was assessed using the 'addition' and 'addition with carry' subtests of the TOBANS (Brigstocke et al., 2016). Children were asked to complete as many arithmetic problems as possible in one minute. In the 'addition' subtask, children were presented with simple addition problems with sums less than ten and in the 'addition with carry' subtask the sums were bigger than ten but smaller than twenty. One point was awarded even if the numeral was written backwards (maximum score $_{\text{addition}}$ = 90; maximum score $_{\text{addition with carry}}$ = 30). This task was administered as a group task.

At Time 4, children were also presented with the 'subtraction' subtask of the TOBANS in addition to the 'addition' and 'addition with carry' subtasks. Similar to 'addition', children were asked to solve as many of the 90 subtraction problems as possible in one minute.

6.2.1.3 Procedure.

The inhibition task, nonsymbolic ANS and arithmetic tasks were part of a comprehensive test battery. Testing was carried out at two consecutive testing points within a 4-month period which were part of a longitudinal study. Testing started when children were in the autumn term of Year One and the second testing took place during the spring term of Year One. Wherever possible, each child was seen by the same experimenter two to four days in a row. The main researcher was assisted by several research assistants from undergraduate psychology classes. They were trained on how to administer the test battery and were given instructions on how to work with young children.

Arithmetic and magnitude comparison task were assessed as group tasks with a ratio of experimenters to children of 1:3. The inhibition task was assessed individually. Testing was carried out tested in a separate room or another quiet place in the school.

Preliminary to the testing, the experimenters attended at least one day in each class so that the children got to know them and felt more comfortable around them. Moreover, the experimenters told each child that they would play games and asked questions such as “How are you?” or “How old are you?”.

All tests included practice items. Concerning feedback, children received only concrete feedback on their performance for practice items and general praise and encouragement throughout the tests.

6.2.2 Results.

The descriptive statistics are shown in Table 6.1. To answer the research question regarding the relationship between inhibition, nonsymbolic ANS and early arithmetic performance, a series of hierarchical multiple regression models were conducted using IBM SPSS Statistics 22. The dependent variable comprised of the composite score of arithmetic subtask raw scores and independent variables were composite scores comprising all nonsymbolic magnitude comparison task and a second independent variable of the composite score of inhibition comprising the raw scores of block one and two of the Head-to-Toe-task.

Furthermore, the same hierarchical regression models were run using MPlus Version 7 and the technique of Cholesky factorisation with phantom factors in a latent variable model (de Jong, 1999) was applied. One advantage of this method is that SEM models use latent variables rather than manifest, observed variables. Latent variables may reduce the dimensionality of data and may impute relationships between unobserved constructs (latent variables) from observable variables. Various measured variables are aggregated to represent an underlying concept. In the SEM path models, the dependent variable was the latent factor of children’s arithmetic scores. The independent variables included the latent factors for magnitude comparison and inhibition.

6.2.2.1 Hierarchical regression models.

6.2.2.1.1 Relationship between inhibition, ANS and arithmetic at Time 3.

The correlation matrix for magnitude comparison, inhibition and arithmetic is shown in Table 6.2. Not surprisingly, all measures related significantly to each other, but magnitude comparison tasks correlated higher with arithmetic than inhibition at Time 3.

Table 6.2

Correlations between ANS, inhibition and arithmetic at Time 3 (n = 76)

	1	2	3
1. Inhibition	---	.327**	.316**
2. Magnitude Comparison		----	.621**
3. Arithmetic			----

*Notes. Pearson product-moment correlation coefficient. Variables entered are composite scores. * p < .05. ** p < .01*

With a composite TOBANS raw score as the dependent variable (arithmetic) and composite scores of magnitude comparison and inhibition task as independent variables, hierarchical regression models were conducted in which magnitude comparison was entered in the first step and inhibition was entered in the second step. As shown in Table 6.3, magnitude comparison significantly predicted performance on arithmetic tasks when entered in step one ($ß$= .621, $t(74) = 6.823$, $p < .001$, $r^2 = .386$), however adding inhibition in step two did not significantly improve the model ($ß$= .126, $t(73) = 1.312$, $p = .194$, $r^2 = .400$, $r_{change}^2 = .014$, $F_{change}(1, 73) = 1.721$, $p = .194$).

A second hierarchical regression model was conducted with the reverse order: The inhibition score was entered in the first step and magnitude comparison was added in the second step (Table 6.4). Inhibition significantly predicted arithmetic when entered in the first step ($ß$= .316, $t(74) = 2.861$, $p = .005$, $r^2 = .100$), but magnitude comparison added significantly to the model when entered in the second step ($ß$= .580, $t(73) = 6.051$, $p < .001$, $r^2 = .400$, $r_{change}^2 = .301$, $F_{change}(1, 73) = 36.610$, $p < .001$). Inhibition was not a significant predictor in the second step.

In other words, inhibition did not significantly explain variance in arithmetic performance at Time 3 once performance on magnitude comparison tasks had been taken into account.

Table 6.3

Hierarchical Regressions for ANS, Inhibition and arithmetic at Time 3 (n = 76) and Time 4 (n = 108)

	Time 3			Time 4		
	B	*SE B*	*ß*	*B*	*SE B*	*ß*
Step 1						
Constant	-3.56	1.78		-8.02	3.47	
Magnitude comparison	.10	.01	.62**	.17	.02	.58**
Step 2						
Constant	-5.22	2.17		-12.91	4.94	
Magnitude comparison	.09	.02	.58**	.16	.03	.54**
Inhibition	.10	.07	.13	.20	.15	.12

*Notes. Time 3t: Step 1 R^2 = .386, Step 2 ΔR^2 = .014. Time 4: Step 1 R^2 = .331, Step 2 ΔR^2 = .012. * $p < .05$. ** $p < .01$.*

Table 6.4

Hierarchical Regressions for Inhibition, ANS and arithmetic at Times 3 and 4

	Time 3			Time 4		
	B	*SE B*	*ß*	*B*	*SE B*	*ß*
Step 1						
Constant	1.85	2.23		-.62	5.32	
Inhibition	.24	.09	.32**	.52	.16	.30**
Step 2						
Constant	-5.22	2.17		-12.91	4.94	
Inhibition	.10	.07	.13	.20	.15	.12
Magnitude comparison	.09	.02	.58**	.16	.03	.54**

*Notes. Time 3t: Step 1 R^2 = .100, Step 2 ΔR^2 = .301. Time 4: Step 1 R^2 = .090, Step 2 ΔR^2 = .253. * p < .05. ** p < .01.*

6.2.2.1.2 Relationship between inhibition, ANS and arithmetic at Time 4.

The correlation matrix for magnitude comparison, inhibition and arithmetic is shown in Table 6.5. All measures related significantly to each other, but magnitude comparison tasks correlated higher with arithmetic than inhibition at Time 4.

The dependent variable was the composite TOBANS raw score (arithmetic) and composite scores of magnitude comparison and inhibition task were the independent variables. Hierarchical regression models, similar to Time 3, were conducted in which magnitude comparison was entered first and inhibition was entered in the second step (Figure 6.3). Magnitude comparison was a significant predictor of children's performance on arithmetic tasks when entered in step one (*ß*= .575, $t(106) = 7.242$, $p < .001$, $r^2 = .331$), but entering inhibition in step two did not significantly improve the model fit (*ß*= .117, $t(105) = 1.388$, $p = .168$, $r^2 = .343$, $r_{change}^2 = .012$, $F_{change}(1, 105) = 1.927$, $p = .168$), as shown in Table 6.4.

Table 6.5

Correlations between ANS, inhibition and arithmetic Time 4 (n = 108)

	1	2	3
1. Inhibition	---	.343**	.301**
2. Magnitude Comparison		----	.575**
3. Arithmetic			----

*Notes. Pearson product-moment correlation coefficient. Variables entered are composite scores. * $p < .05$. ** $p < .01$*

As shown in Table 6.4, the second hierarchical regression model was conducted with the reverse order: The inhibition score was entered in the first step and magnitude comparison was added in the second step. This time, inhibition did significantly predicted arithmetic when entered in the first step (*ß*= .301, $t(106) = 3.246$, $p = .002$, $r^2 = .090$), but magnitude comparison added significantly to the model when entered in the second step (*ß*= .535, $t(105) = 6.355$, $p < .001$, $r^2 = .343$, $r_{change}^2 = .253$, $F_{change}(1, 105) = 40.385$, $p < .001$). Inhibition was not a significant predictor in the second step.

The findings confirmed the results form Time 3 that inhibition did not significantly explain variance in arithmetic performance once performance on magnitude comparison tasks had been taken into account.

6.2.2.2. Hierarchical regressions using structural equation modelling.

6.2.2.2.1 Relationship between inhibition, ANS and arithmetic at Time 3.

To further strengthen the findings of the hierarchical regression models using IBM SPSS Statistics 22, the same hierarchical regression models were conducted using MPlus Version 7 and the technique of Cholesky factorisation with phantom factors in a latent variable model (de Jong, 1999) was applied. In the SEM path models, the dependent variable was the latent factor of children’s arithmetic scores. The independent variables included the latent factors for magnitude comparison and inhibition task. In Cholesky factorisation, the individual steps of the hierarchical regression were coded as phantom latent factors which were then regressed onto arithmetic outcome scores.

In the first hierarchical regression model (Figure 6.8), the phantom latent factor magnitude comparison was entered first and the phantom latent factor including magnitude comparison and inhibition was entered in the second step. Both phantom latent factors were then regressed onto arithmetic performance at Time 3. The path model provided an acceptable fit to the data, $\chi 2$ (71) = 97.323, *p* = .021, *RMSEA* = .057 (90% CI = .023 - .083), *CFI* = .976, *SRMR* = .043, explaining 46.4% of variance. Both phantom latent factors significantly predicted arithmetic meaning that step one, magnitude comparison, as well as step two, magnitude comparison and inhibition, were significant.

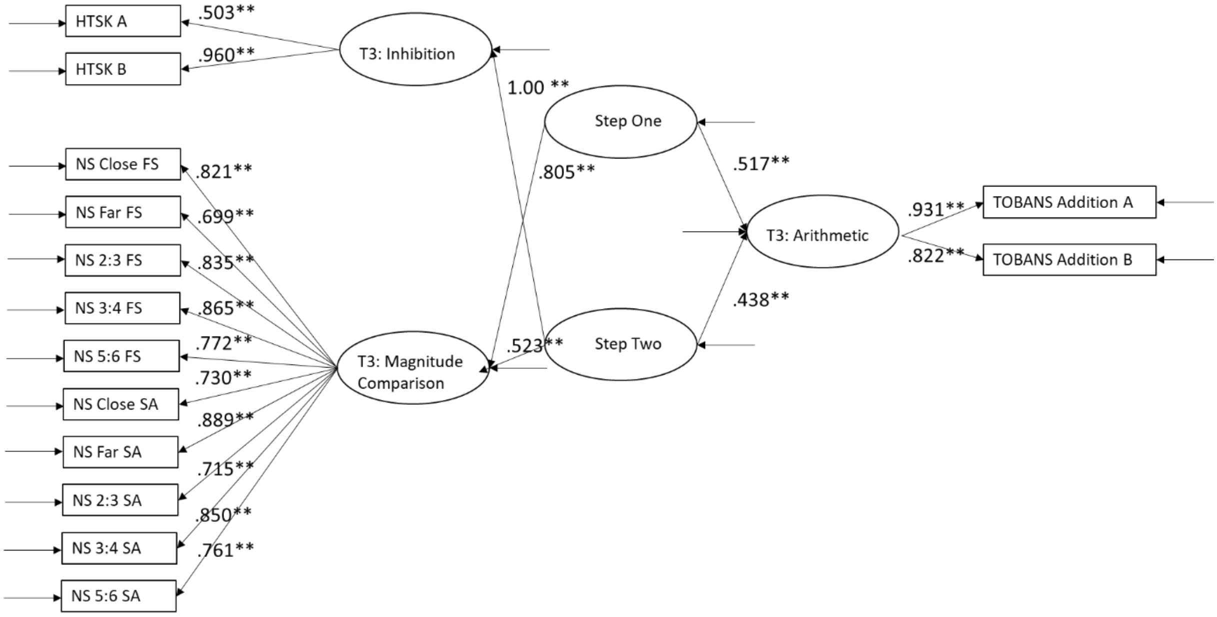


Figure 6.8. Hierarchical SEM regression model of arithmetic at Time 3. Step 1: Magnitude Comparison. Step 2: Magnitude Comparison and Inhibition. * p < .05. ** p < .01.

The second hierarchical regression model (Figure 6.9) was conducted with the reverse order: The inhibition score was coded as the first phantom latent factor (step one) and inhibition and magnitude comparison coded as the second phantom latent factor (step two). The path model provided an acceptable fit to the data, $\chi 2$ (71) = 97.323, *p* = .021, *RMSEA* = .057 (90% CI = .023 - .083), *CFI* = .976, *SRMR* = .043, explaining 46.4% of variance. The phantom latent factor for step two (inhibition and magnitude comparison) was the only unique predictor of arithmetic performance. The phantom latent factor including only inhibition did not significantly contributed to explaining the variance.

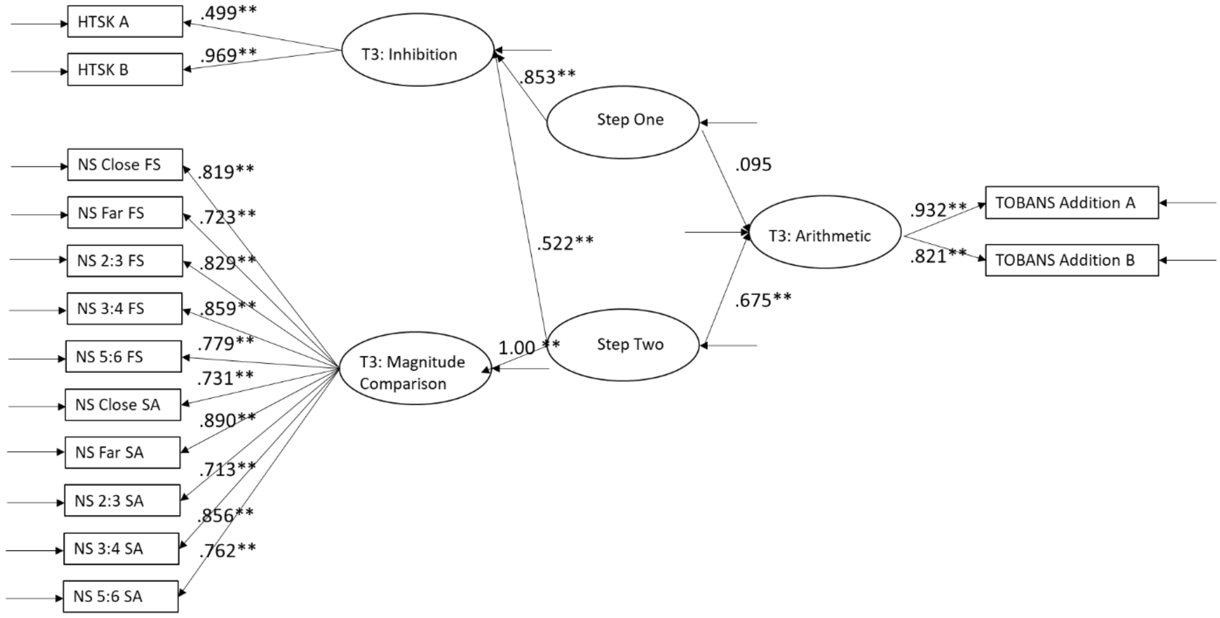


Figure 6.9. Hierarchical SEM regression model of arithmetic at Time 3. Step 1: Inhibition. Step 2: Inhibition and Magnitude Comparison. * p < .05. ** p < .01.

In other words, the results of the SEM path models confirmed the findings of the SPSS hierarchical regressions that inhibition did not significantly explain variance in arithmetic performance at Time 3 once performance on magnitude comparison tasks had been taken into account.

6.2.2.2.2 Relationship between inhibition, ANS and arithmetic at Time 4.

Similar to Time 3, the first hierarchical regression model (Figure 6.10) included the phantom latent factor magnitude comparison at Time 4 which was entered first and the phantom latent factor including magnitude comparison and inhibition which was entered in the second step. Both phantom latent factors were then regressed onto arithmetic performance at Time 4. The path model provided a moderate fit to the data, $\chi 2$ (98) = 142.120, *p* = .002, *RMSEA* = .063 (90% CI = .038 - .085), *CFI* = .961, *SRMR* = .058, explaining 39.3% of variance. Again, both phantom latent factors were significantly predicting arithmetic suggesting that step one, magnitude comparison, as well as step two, magnitude comparison and inhibition, were significantly predicting arithmetic score.

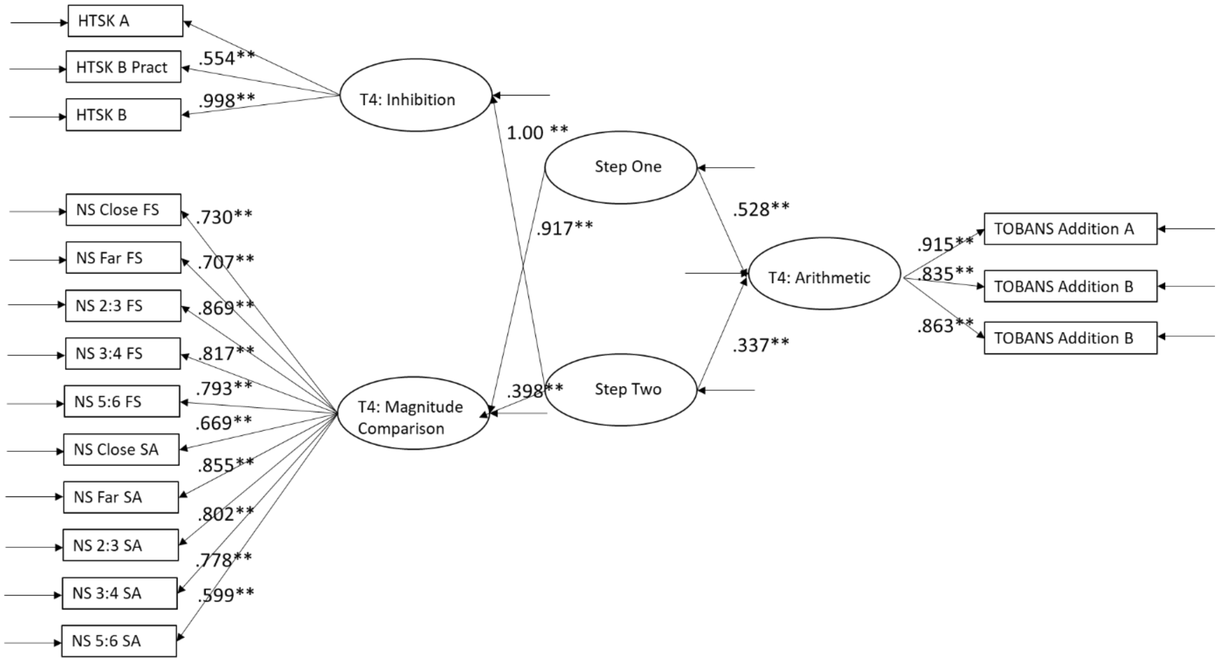


Figure 6.10. Hierarchical SEM regression model of arithmetic at Time 4. Step 1: Magnitude Comparison. Step 2: Magnitude Comparison and Inhibition. * p < .05. ** p < .01.

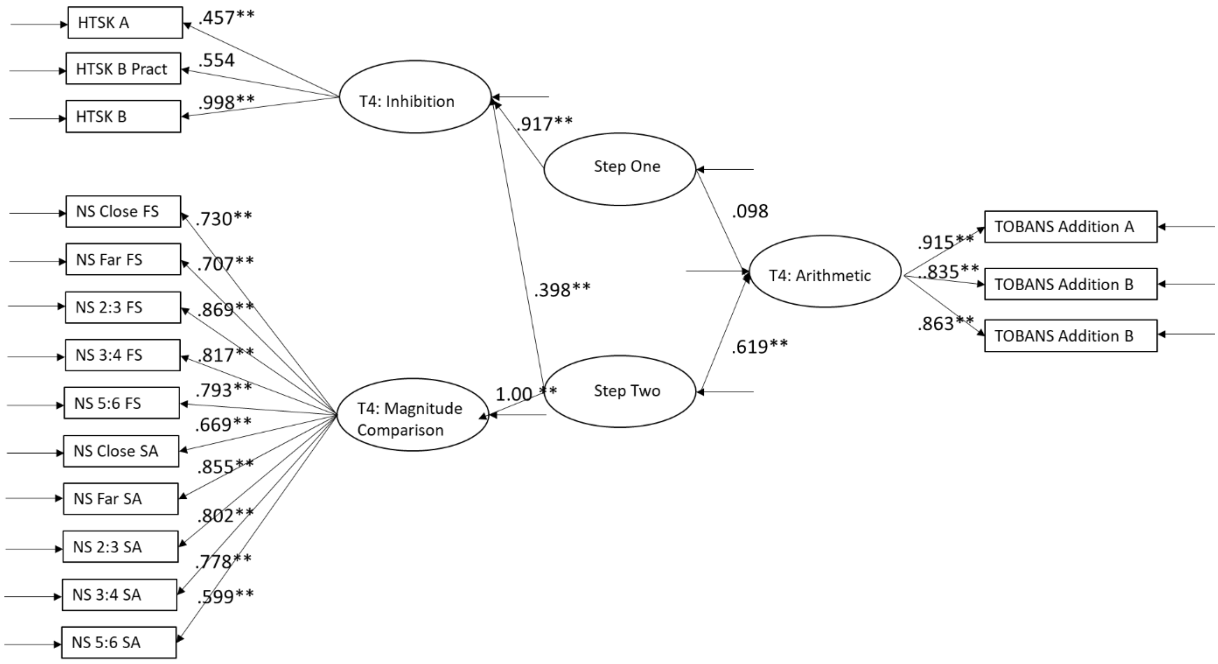


Figure 6.11. Hierarchical SEM regression model of arithmetic at Time 4. Step 1: Inhibition. Step 2: Inhibition and Magnitude Comparison. * p < .05. ** p < .01.

The second hierarchical regression model (Figure 6.11) was conducted in reverse order: The inhibition score was coded as the first phantom latent factor (step one) and inhibition and magnitude comparison coded as the second phantom latent factor (step two). The path model provided an acceptable fit to the data, $\chi2$ (98) = 142.120, *p* = .002, *RMSEA* = .063 (90% CI = .038 - .085), *CFI* = .961, *SRMR* = .058, explaining 39.3% of variance. The phantom latent factor for step two (inhibition and magnitude comparison) was the only unique predictor of arithmetic performance. The phantom latent factor including only inhibition did not contribute significantly.

These findings confirmed the Time 3 results and the Time 4 results of the SPSS hierarchical regression models that inhibition did not significantly explain variance in arithmetic performance at Time 4 once performance on magnitude comparison tasks had been taken into account.

6.2.3 Conclusion.

Contrary to the findings of Gilmore et al. (2013) showing that children's inhibitory control predicted arithmetic after controlling for performance on dot comparison, this study showed the reverse pattern. It is worth mentioning that the sample in Gilmore et al. (2013) were older and, despite the broad age range (seven to ten years), the analyses did not control for age. Also, this study assessed a different tsk of a GoNoGo inhibition task than in Gilmore et al. (2013) using the NEPS-II inhibition subtask (Korkman, Kirk, and Kemp, 1998), a GoNoGo test. Both inhibition tasks are a GoNoGo inhibition test but it may be possible that they measure different aspects of inhibition, thus causing the contrasting results.

The finding of the current study was that children's performance on inhibition tasks did not explain variance of arithmetic scores once performance on nonsymbolic magnitude comparison has been accounted for. Both conventional hierarchical regression models and SEM phantom latent factor regressions confirmed that the ANS is more important in the development of early arithmetic. It must be mentioned, that the model fit indices of the path models were moderate at best. Unique variance per predictor was low indicating that shared variance is substantial. Neither inhibitory control nor nonsymbolic magnitude comparison may play the most important role in predicting arithmetic. These results support the findings from the longitudinal prediction in Chapter 5 that transcoding skills, children's understanding of the Arabic numeral system, was the only stable longitudinal precursor of early arithmetic skills.

Chapter 7. Approximate Arithmetic Performance

The main focus of this chapter was to examine the developmental relation between symbolic and nonsymbolic approximate arithmetic, and the developmental relation between approximate and exact arithmetic. The chapter will then further explore longitudinal predictors of symbolic and nonsymbolic approximate arithmetic (Barth et al., 2005; 2006; Gilmore et al., 2007). Also, I will compare the performance of the current sample with the results from Gilmore et al. (2007), focusing on accuracy, ratio effects and differences in performance between symbolic and nonsymbolic approximate arithmetic. Gilmore et al. (2007) showed that pre-school children, before formal training in arithmetic, are capable of performing approximate arithmetic, based on double digit numbers, with accuracy above chance; Gilmore proposed that this ability is based on nonsymbolic approximate representations. Halberda and Feigenson (2008) showed that the acuity of the nonsymbolic ANS increases between three and six years of age and may not reach adult-like levels of performance until early adolescence. Gilmore et al. (2007) referred to three signature properties of nonsymbolic number representation: (1) Performance on comparison, addition and subtraction tasks are subject to ratio limits. (2) Addition is as accurate as performance on comparison problems. (3) Subtraction is less accurate then comparison (Gilmore et al., 2007, p. 590). The ratio limit in particular is of interest to this study, as research shows that accuracy on nonsymbolic number representations falls as the ratios to be compared approach one (Barth et al., 2005; 2006).

Furthermore, this chapters aims to identify the structure of the relation between symbolic and nonsymbolic approximate arithmetic tasks. As seen in Chapter 3, there is evidence that the structure of the relation between magnitude comparison tasks shift over time from a two-factor (symbolic and nonsymbolic comparison) towards a general comparison factor in Year One (6 years of age). Pre-school children still have an immature knowledge of the Arabic numeral system and may hence rely heavily on general cognitive resources as well as magnitude estimation processes to solve symbolic comparison tasks. However, after having mastered the Arabic numeral system, solving symbolic comparison may draw directly on estimation processes similar to those involved in the nonsymbolic comparison tasks. Arguably, approximate arithmetic may behave similarly with two distinct factors for pre-school children (symbolic versus nonsymbolic approximate arithmetic) which may shift towards a general approximate arithmetic factor. Thus, detailed confirmatory factor analyses were conducted examining this relationship.

7.1 Methods.

7.1.1 Participants.

The same participants were used as described in Chapter 2 (p. 42)

7.1.2 Materials.

Children were assessed on the following measures.

7.1.2.1 Baseline Prediction Model assessed at Time 1.

The following tasks were administered individually to the four nursery classes in the summer term of the nursery age (4 years of age): Nonverbal intelligence (Raven's CPM; Raven et al., (1993), grammatical ability (TROG-2; Bishop, 2003), vocabulary (BPVS - III; Dunn et al., 2010), specific math-related language ability (TRC), transcoding (Number Identification, Number Writing and Reading Arabic numerals), rote counting and magnitude comparison tasks.

7.1.2.2 Measures taken at Time 3.

Magnitude Comparison. Various symbolic and nonsymbolic comparison tasks were created for the study, based on those by Göbel et al. (2014). Each comparison pair was presented on a single page. Children were given one point for every correct comparison with a maximum score of 16 for each version and 160 overall (for more details see Chapter 3, pp.50).

Arithmetic Skills. *Fluency.* Children's speeded arithmetic skills (fluency) was assessed using the 'addition' and 'addition with carry' subtests of the TOBANS (Brigstocke et al., 2016). Children were asked to complete as many arithmetic problems as possible in one minute. In the 'addition' subtask, children were presented with simple addition problems with sums less than ten and in the 'addition with carry' subtask the sums were bigger than ten but smaller than twenty. One point was awarded even if the numeral was written backwards (maximum score $_{\text{addition}}$ = 90; maximum score $_{\text{addition with carry}}$ = 30). This task was administered as a group task.

Approximate arithmetic. Symbolic and nonsymbolic approximate arithmetic problems were assessed based on those typically used in literature (Barth et al., 2005;

2006; Gilmore et al., 2007). Children were given the problems both verbally and visually on a computer screen. Initially, a female character called Sarah appeared on the screen with a bag while the experimenter stated that “Sarah has fifteen candies (or “a bag full of that many marbles” in the nonsymbolic part). The Arabic numeral or the appropriate number of dots corresponding to the trial were displayed in the bag. On the next screen, a second bag appeared above the same character and the experimenter stated that “Sarah gets nineteen more candies” (or “she gets that many more marbles”). Again, the Arabic numeral or dots were displayed in the bag. A plus sign connected the two bags only in the symbolic version. On the last screen, children could see Sarah and her two bags and a second character called John. John had a different coloured bag and the experimenter stated that “John has fifty-one candies” (or “a bag full of that many marbles”). Similar to Sarah, the correct Arabic numeral or number of dots was displayed in the bag (Figure 7.1). Finally, the experimenter asked the child who they think has more candies (marbles). See Appendix E for the complete list of trials (Appendix 14).

In the nonsymbolic condition (Appendix 15), children were asked not to count the dots but rather estimate who they think has more. If a child attempted to count the dots, the experimenter reminded the child not to count the dots. The pictures were displayed shortly on the screen to further discourage counting strategies.

First, children had to solve the symbolic problems followed by the nonsymbolic problems. Each condition comprised of 24 problems of larger numbers in the range 5 to 58 divided into three ratios – 4:7; 4:6; 4:5 – eight trials per ratio. The sum was greater than the comparison number on half the trials. The same comparisons were used for symbolic and nonsymbolic conditions. The task was administered in a one-on-one setting and one point was awarded for each correct answer.

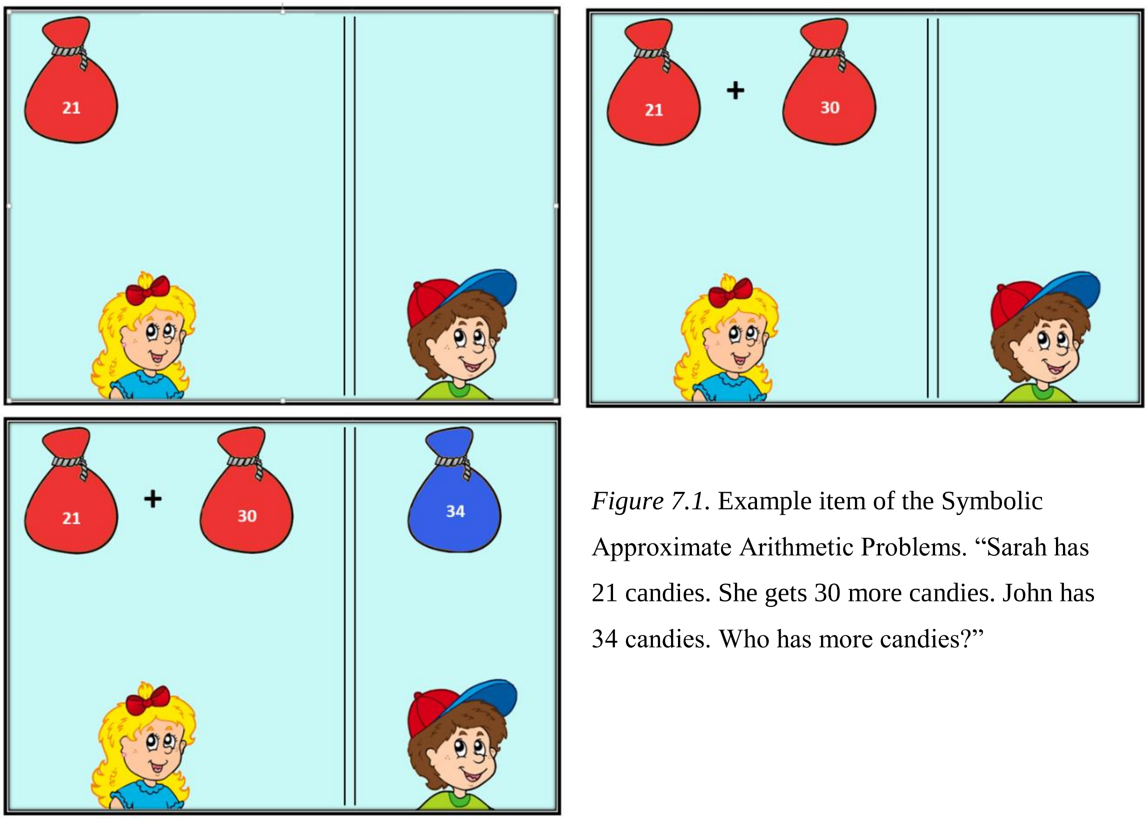


Figure 7.1. Example item of the Symbolic Approximate Arithmetic Problems. “Sarah has 21 candies. She gets 30 more candies. John has 34 candies. Who has more candies?”

7.1.2.3 Measures taken at Time 4.

Magnitude Comparison. The same tasks as at Time 3 were used.

Arithmetic. *Fluency.* In addition to the same TOBANS tasks at Time 3 (addition and addition with carry), children were also presented with the ‘subtraction’ subtask (Brigstocke et al., 2016). Similar to addition, children were asked to solve as many of the 90 subtraction problems as possible in one minute.

Approximate arithmetic. The same tasks as at Time 3 were used.

7.1.2.4 Measures taken at Time 5.

Magnitude Comparison. The same tasks as at Time 3 were used.

Arithmetic. *Fluency.* The same tasks as at Time 4 were used.

Accuracy. Children’s basic arithmetic skills were assessed using the Numerical Operations subtest of the second edition of the Wechsler Individual Achievement Test (WIAT-II; Wechsler, 2005). The first six items (identifying and writing Arabic numerals) were excluded because we were only interested in a more conventional measure of arithmetic. The test was executed according to the manual and children were allowed to complete the task in their own time (maximum score = 25).

Approximate arithmetic. The same tasks as at Time 3 were used.

7.1.3 Procedure.

The ANS and arithmetic tasks were part of a comprehensive test battery. Testing was carried out at three consecutive testing points within a 10-month period which were part of a longitudinal study. Testing started when children were in the autumn term of Year One, the second testing took place during the spring term of Year One and the last testing session was carried out in the summer term of Year One. Wherever possible, each child was seen by the same experimenter two to four days in a row. The main researcher was assisted by several research assistants from undergraduate psychology classes. They were trained on how to administer the test battery and were given instructions on how to work with young children.

TOBANS, WIAT and magnitude comparison tasks were assessed as group tasks with a ratio of experimenters to children of 1:3. The approximate arithmetic task was assessed individually on a laptop or tablet. Children were discouraged to use counting strategies to solve the problems Testing was carried out tested in a separate room or another quiet place in the school. Preliminary to the testing, the experimenters attended at least one day in each class so that the children got to know them and felt more comfortable around them. Moreover, the experimenters told each child that they would play games and asked questions such as “How are you?” or “How old are you?”. All unstandardized tests included practice items. Concerning feedback, children received only concrete feedback on their performance for practice items and general praise and encouragement throughout the tests.

7.2 Results.

To address research questions concerning approximate arithmetic, descriptive statistics were analysed first, followed by the analysis of ratio effects over time and correlations between the approximate arithmetic and TOBANs and WIAT arithmetic subtasks. A set of CFAs path models were estimated with Mplus Version 7 (Muthén and Muthén, 2013) to examine whether the relationship between symbolic and nonsymbolic approximate arithmetic changes from a two-factor model towards a unitary model, similar to the magnitude comparison task, with the latent factor symbolic approximate arithmetic and nonsymbolic approximate arithmetic.. Last, SEM path models were conducted investigating the longitudinal prediction of approximate arithmetic, with the latent factor arithmetic (consisting of TOBANS subtasks at Time 3 and TOBANS and WIAT at Time4) as the dependent variable and the Time 1 baseline model as independent variables. To visually simplify the path models, the coefficients of the relations between the factors are not presented but can be found in Appendices 23 and 24.

Table 7.1

Mean and standard deviations of predictor and criterion measures from all testing sessions

		Time 1	Time 3	Time 4	Time 5
		M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
Nonverbal IQ	Raven's CPM	6.45 (1.57)			
Language Comprehension	TROG-2	3.15 (2.63)*			
Vocabulary	BPVS-III	58.26 (16.77)*			
Math-related Language	TRC	5.90 (2.00)*			
Numerical Knowledge	Number Writing	6.86 (5.79)*			
	Number Reading	8.02 (2.67) [45]*			
	Number Identification	7.26 (2.60)*			
	Rote Counting	14.78 (12.84)*			
Magnitude Comparison	Digit Close	10.11 (3.21) [7]			
	Digit Far	10.73 (3.59) [7]			
	NS FS Close	10.26 (2.16) [1]			

	NS FS Far	12.87 (2.57) [17]			
	NS FS 3:4				
	NS FS 5:6	10.76 (2.72) [4]			
	NS SA Close	10.58 (2.29) [3]			
	NS SA Far	10.29 (2.28) [1]			
	NS SA 2:3	13.24 (2.43) [23]			
	NS SA 3:4	11.42 (2.42) [5]			
		10.81 (2.03) [2]*			
Arithmetic	Addition Tasks	6.33 (3.21) [17]*	A: 5.23 (2.51) [3]		
			B: 5.15 (2.55) [8]*		
	TOBANS				
	Addition		6.23 (4.55)	8.36 (5.09)	12.74 (8.66)
	Addition w/ carry		1.75 (2.20)	2.56 (2.74)	5.07 (5.01) [1]
	Subtraction			5.30 (4.12)	8.44 (5.10)
	Approximate		Symbolic: 14.30 (.3.54)	Symbolic: 15.87 (3.74) [4]	Symbolic: 18.37 (3.66) [13]
	Arithmetic		NS: 15.70 (3.41)*	NS: 18.02 (3.09) [1]*	NS: 19.16 (2.76) [5]*
	WIAT				4.00 (2.27)

Notes. *M = mean age. SD = standard deviation* * *individually administered tasks. The number of children scoring maximum level are shown in square brackets. All scores are presented as raw scores. For the Magnitude Comparison Tasks: NS = nonsymbolic. FS = fixed size trials. SA = surface-area matched trials.*

7.2.1 Descriptive Statistics.

The descriptive analysis of all measures are shown in Table 7.1. Descriptive statistics were conducted using IBM SPSS Statistics 22. Performance on the approximate arithmetic tasks increased over time and nonsymbolic approximate arithmetic was more accurate than symbolic approximate arithmetic. A ceiling effect was present at Time 5, with thirteen children reaching the maximum score on the symbolic approximate arithmetic and five children scoring at ceiling level on the nonsymbolic approximate arithmetic.

Children answered 59.17% of the symbolic items correctly and 65.00% of nonsymbolic items at Time 3. On average, children did perform significantly better on nonsymbolic trials (M = 15.70, SD = 3.41) than symbolic trials (M = 14.30, SD = 3.54), $t(114) = 4.644$, $p < .001$. The correlation between symbolic and nonsymbolic conditions was highly significant, $r = .557$, $p < .001$.

Children's performance increased in accuracy, with 75.07% of correctly answered nonsymbolic trials compared to 68.58% of correctly answered symbolic trials at Time 4. Children performed significantly better on nonsymbolic trials (M = 18.02, SD = 3.09) than symbolic trials (M = 16.59, SD = 3.74), $t(112) = 4.812$, $p < .001$. The correlation between symbolic and nonsymbolic conditions was highly significant, with $r = .491$, $p < .001$.

Accuracy at Time 5 increased even further with 79.60% of accuracy on nonsymbolic trials and 75.79% of accuracy on symbolic trials. The performance on nonsymbolic trials (M = 19.16, SD = 2.76) significantly different from the performance on symbolic trials (M =18.37, SD = 3.67), $t(115) = 2.917$, $p = .004$. Symbolic and nonsymbolic approximate arithmetic correlated highly, $r = .520$, $p < .001$.

One-sample two-tailed t-tests comparing performance on symbolic and nonsymbolic approximate arithmetic with chance level (50%) showed that children significantly performed above chance level on all symbolic and nonsymbolic arithmetic problems (see Table 7.2 for individual t-test statistics).

Table 7.2

Children's Performance on Symbolic and Nonsymbolic Approximate Arithmetic compared to chance level at Times 3, 4 and 5

	Symbolic Approximate Arithmetic				Nonsymbolic Approximate Arithmetic			
	t value	Degrees of Freedom	*p* value	*M (SD)*	*t* value	Degrees of Freedom	*p* value	*M (SD)*
Time 3	6.755	1, 114	<.001	14.20 (3.49)	11.452	1, 114	<.001	15.60 (3.37)
Time 4	12.853	1, 112	<.001	16.46 (3.69)	21.054	1, 112	<.001	18.02 (3.04)
Time 5	17.453	1, 115	<.001	18.19 (3.82)	26.835	1, 115	<.001	19.10 (2.85)

Notes. M = mean. SD = standard deviation. One-sample two-tailed t-Tests comparing to chance level (50%).

7.2.2 Ratio effects.

To test for ratio effects, children's performance at each of the three ratios was compared for each form of presentation (symbolic versus nonsymbolic). Simple 3 (ratio: 4:5, 4:6, 4:7) by 2 (presentation: symbolic versus nonsymbolic) ANOVAs were conducted for each testing point. Analysis of the ratios at Time 3 showed a main effect for ratio ($F(2,228) = 52.715$, $p < .001$, $\eta_p^2 = .316$) and presentation ($F(1,114) = 22.523$, $p < .001$, $\eta_p^2 = .165$), but no significant interaction. Children showed greater performance on nonsymbolic trials ($M = 5.21$, $SD = .11$) than symbolic ($M = 4.728$, $SD = .11$), and performed significantly better on 4:7 ratio ($M = 5.42$, $SD = .12$), followed by 4:6 ratio ($M = 5.01$, $SD = .11$) and 4:5 ratio ($M = 4.47$, $SD = .09$).

At Time 4, main effects for ratio ($F(2,230) = 54.246$, $p < .001$, $\eta_p^2 = .321$) and presentation ($F(1,115) = 8.649$, $p = .004$, $\eta_p^2 = .070$) were present. Children showed greater performance on nonsymbolic trials ($M = 6.01$, $SD = .10$) than symbolic ($M = 5.487$, $SD = .12$). Ratio 4:7 was the easiest ($M = 6.27$, $SD = .11$)

compared to 4:6 (*M* = 5.84, *SD* = .11) and 4:5 (*M* = 5.13, *SD* = .10; all ratio comparisons were significant).

Analysis of Time 5 data revealed the same findings as in previous time points, with significant main effects for ratio ($F(2,232) = 54.246$, $p < .001$, $\eta_p^2 = .321$) and presentation ($F(1,116) = 8.649$, $p = .041$, $\eta_p^2 = .07$), but no significant interaction. Performance on nonsymbolic stimuli (M = 6.37, SD = .09) was greater than symbolic (M = 6.06, SD = .12). Inspection of children's performance on ratio trials showed greater performance on 4:7 (*M* = 6.69, *SD* = .10) ratios compared to 4:6 (*M* = 6.28, *SD* = .11) and compared to 4:5 (*M* = 5.67, *SD* = .11).

7.2.3 Exploration of the structure of approximate arithmetic.

Next, the relation between the approximate arithmetic tasks and exact arithmetic tasks was assessed. Exact arithmetic was defined as children's performance on composite scores. First, the correlations between the composite scores of exact arithmetic, including both TOBANS and WIAT, and symbolic as well as nonsymbolic approximate arithmetic were conducted (Figure 7.3). The composite score of symbolic approximate arithmetic comprised of the scores of the three ratios 4:5, 4:6 and 4:7 and separately, for nonsymbolic tasks. At Time 3, the composite of the two TOBANS addition subtasks scores were used whereas Times 4 and 5 consisted of composite scores of the TOBANS addition and subtraction subtasks and WIAT.

Table 7.3

Correlation matrix between Symbolic and Nonsymbolic Approximate Arithmetic and Exact Arithmetic at Times 3, 4 and 5

	Exact Arithmetic		
	Time 3	Time 4	Time 5
Symbolic approximate arithmetic	.499**	.530**	.389**
Nonsymbolic approximate arithmetic	.392**	.275**	.260**

*Notes. Pearson product-moment correlation coefficient. * p < .05. ** p < .01.*

Both the symbolic and nonsymbolic approximate arithmetic skills significantly correlated with the TOBANS at Time 3. The correlation between TOBANS and symbolic approximate arithmetic was higher (r = .499) than nonsymbolic (r = .392). Similarly, symbolic was highly related to TOBANS at Time 4 (r = .530). The correlation between TOBANS and nonsymbolic approximate arithmetic at Time 4 was significant (r = .275) but much lower than symbolic. The correlations at Time 5 were overall lower than the other two testing points but still significant. The correlation to symbolic approximate arithmetic was again higher (r = .389) compared to nonsymbolic approximate arithmetic (r = .260).

Similar to the analysis of the structure of the magnitude comparison (see Chapter 3), a series of confirmatory factor analyses were conducted to investigate the relation between the two approximate arithmetic tasks. The CFAs examined in what way symbolic and nonsymbolic approximate arithmetic latent factors are related by comparing a one-factor (general approximate arithmetic ability) and a two-factor model (symbolic and nonsymbolic approximate arithmetic). Furthermore, it was investigated whether this relationship changes over time and if the structure switches from a two-factor towards a unitary factor model as was found in the case of magnitude comparison.

7.2.3.1 Time 3. The first set of CFAs examined the nature of approximate arithmetic tasks at Time 3. All tasks loaded significantly on the single factor approximate arithmetic CFA (Figure 7.2). The model did not provide an acceptable fit to the data, χ^2 (9) = 38.328, p < .001, *RMSEA* = .168 (90% CI = .116 - .225), *CFI* = .872, *SRMR* = .066, suggesting that a single factor is not sufficient and a better model would involve at least the two factors symbolic and nonsymbolic approximate arithmetic. Figure 7.2 shows the two-factor model which provided an excellent fit to the data, χ^2 (8) = 13.712, p = .090, *RMSEA* = .079 (90% CI = .000 - .148), *CFI* = .975, *SRMR* = .007. A chi-squared difference test confirmed that this model fits the data significantly better than the unitary model (χ^2_{diff} (1) = 24.616, p < .001).

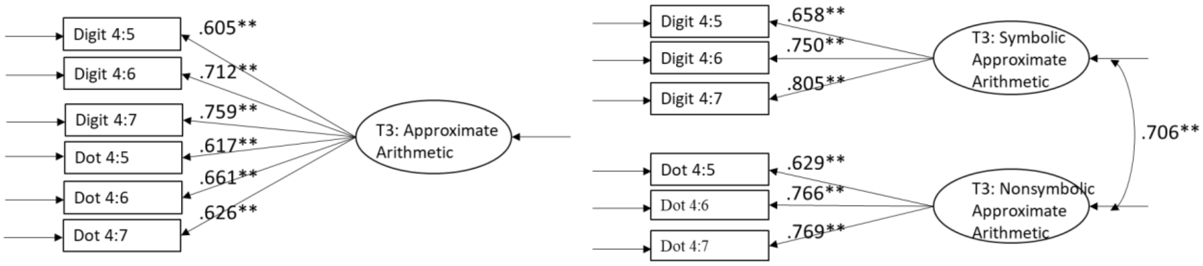


Figure 7.2. One factor (left side) and two factor (right side) CFA of symbolic and nonsymbolic approximate arithmetic tasks (Time 3). $p < .001$. * $p < .05$, ** $p < .01$

7.2.3.2 Time 4. The second set of CFAs (Figure 7.3) assessed the relationship of approximate arithmetic at Time 4. The first model presented the single factor model. Although all variables loaded significantly onto the hypothesised general approximate arithmetic factor, it is worth mentioning that the nonsymbolic ratio 4:5 had a low loading. The model did not provide an adequate fit to the data, χ^2 (9) = 40.111, $p < .001$, *RMSEA* = .175 (90% CI = .122 - .232), *CFI* = .854, *SRMR* = .078. Contrary, the two-factor model provided an excellent fit to the data χ^2 (8) = 11.216, p = .190, *RMSEA* = .060 (90% CI = .000 - .134), *CFI* = .985, *SRMR* = .038, which was significantly better than the single-factor model (χ^2_{diff} (1) = 28.895, $p < .001$).

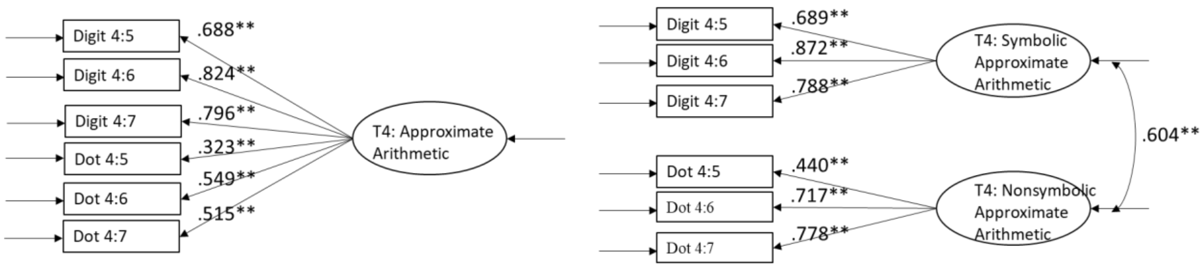


Figure 7.3. One factor (left side) and two factor (right side) CFA of symbolic and nonsymbolic approximate arithmetic tasks (Time 4). $p < .001$. * $p < .05$, ** $p < .01$

7.2.3.3 Time 5. The last set of CFAs investigated the structure of the approximate arithmetic tasks at Time 5 (Figure 7.4). The single-factor model provided a weak fit to the data, χ^2 (9) = 20.253, p = .016, *RMSEA* = .104 (90% CI = .042 - .165), *CFI* = .935, *SRMR* = .053 compared to the two-factor model which provided an excellent fit to the data, χ^2 (8) = 8.946, p = .347, *RMSEA* = .032 (90% CI = .000 - .116), *CFI* = .995, *SRMR* = .032. However, the chi-squared difference test confirmed that the two-factor model fitted the data significantly better than the single-factor model (χ^2_{diff} (1) = 11.307, $p < .001$).

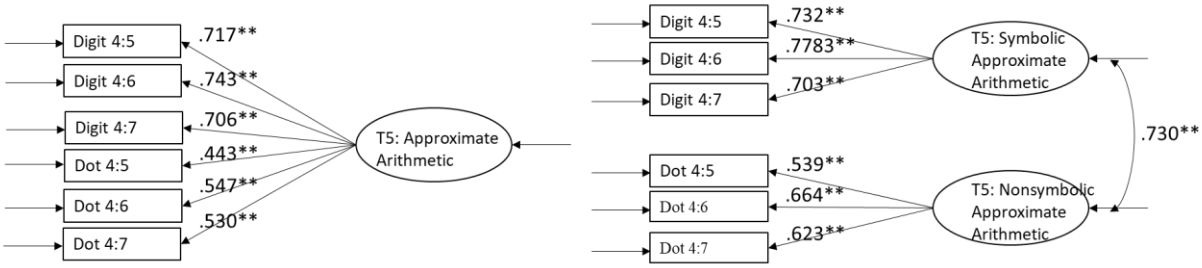


Figure 7.4. One factor (left side) and two factor (right side) CFA of symbolic and nonsymbolic approximate arithmetic tasks (Time 5). $p < .001$. * $p < .05$, ** $p < .01$

7.2.4 Predicting approximate arithmetic using Time 1 baseline model.

To assess the longitudinal prediction of approximate arithmetic, the Time 1 base model (latent independent variables: nonverbal intelligence, general language comprehension, math-related language and transcoding, counting skills as well as the two magnitude comparison constructs as used in Chapter 4, pp. 76-79) was regressed onto the dependent latent factor symbolic and, separately, nonsymbolic approximate arithmetic at Times 3, 4 and 5. Latent variables assessed by only one indicator were pre-specified with an error reflecting the reliability of the variable calculated on the sample to minimise distortions caused by measurement errors.

The path model depicted in Figure 7.5 shows the longitudinal predictors of symbolic approximate arithmetic at Time 3. The model provided an excellent fit to the data, χ^2 (100) = 117.541, p = .111, *RMSEA* = .035 (90% CI = .000 - .059), *CFI* = .959, *SRMR* = .074, and symbolic magnitude comparison at Time 1 was the only unique predictor of symbolic approximate arithmetic at Time 3 (41.3% of variance explained). Likewise, nonsymbolic approximate arithmetic at Time 3 was uniquely predicted by children's performance on the symbolic comparison at Time 1, χ^2 (100) = 112.078, p = .193, *RMSEA* = .029 (90% CI = .000 - .055), *CFI* = .970, *SRMR* = .067 (Figure 7.6). The model only explained 25.1% of variance.

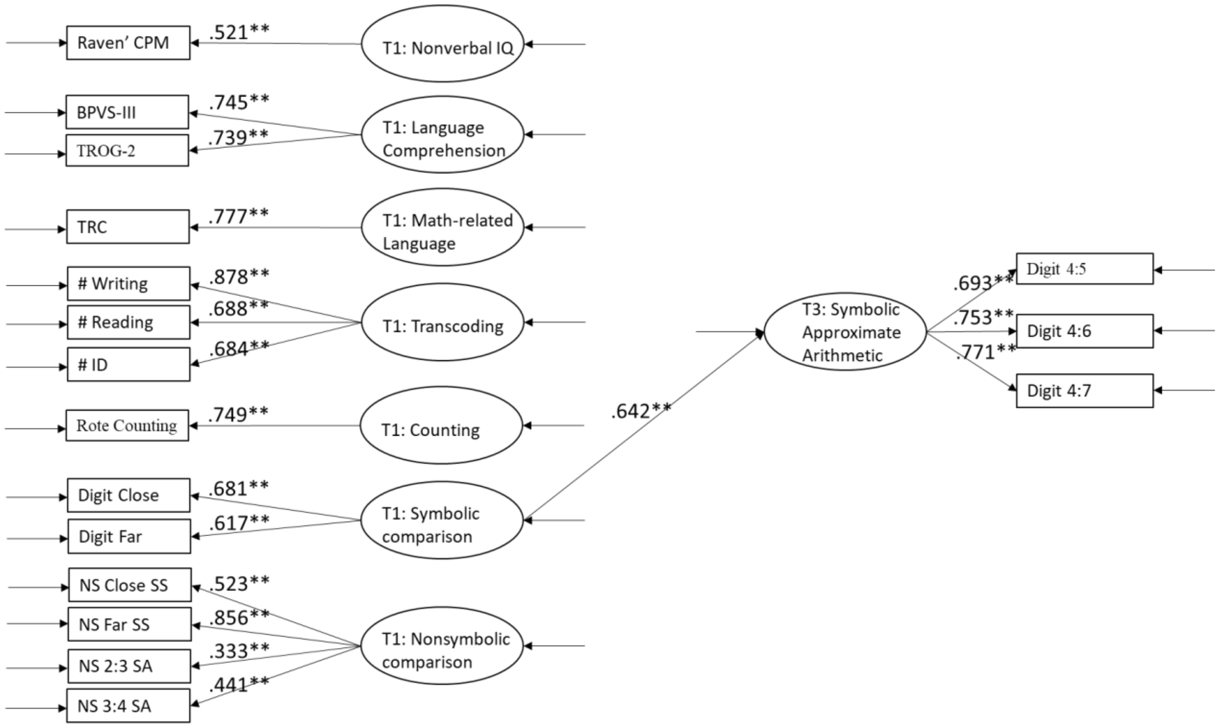


Figure 7.5. Prediction of symbolic approximate arithmetic at Time 3 by Time 1 base model. * p < .05. ** p < .01.

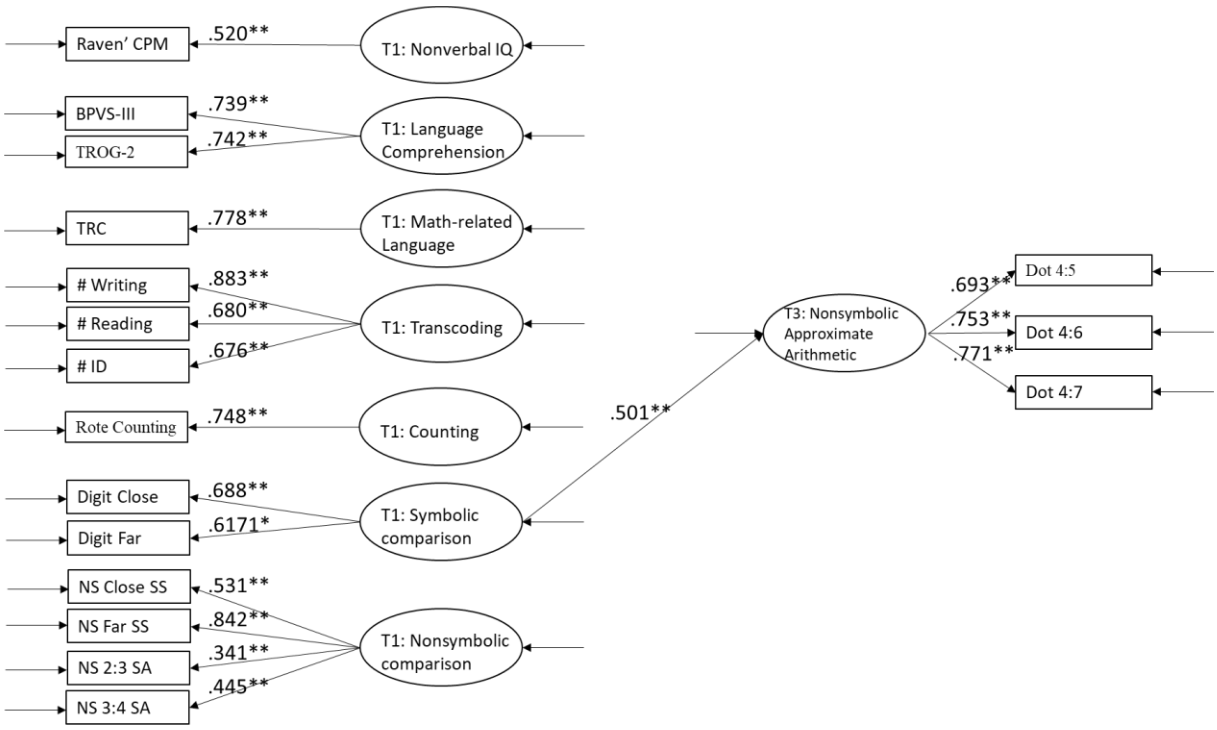


Figure 7.6. Prediction of nonsymbolic approximate arithmetic at Time 3 by Time 1 base model. * p < .05. ** p < .01.

At Time 4, symbolic approximate arithmetic scores were uniquely predicted by symbolic approximate arithmetic at Time 3 (autoregressor) and transcoding at Time 1. The model provided an excellent fit to the data, χ^2 (149) = 170.826, p = .107, *RMSEA* = .032 (90% CI = .000 - .052), *CFI* = .964, *SRMR* = .074, (shown in Figure

7.7). The autoregressor was predicted by symbolic magnitude comparison at Time 1 (41.4% of variance of symbolic approximate arithmetic at Time 3 and 54.1% of variance of arithmetic at Time 4 was explained). Children's nonsymbolic approximate arithmetic, shown in Figure 7.8, fitted the data acceptably, χ^2 (149) = 180.147, p = .042, *RMSEA* = .038 (90% CI = .000 - .057), *CFI* = .937, *SRMR* = .080. Interestingly, both nonverbal intelligence (Time 1) and the autoregressor (predicted by symbolic magnitude comparison at Time 1), were unique predictors of nonsymbolic approximate arithmetic at Time 4. Nonverbal intelligence was the stronger predictor (symbolic magnitude comparison at Time 1 explained 22.8% of variance of nonsymbolic approximate arithmetic at Time 3, and the autoregressor at Time 3 and nonverbal ability at Time 1 explained 51.3% of variance at Time 4).

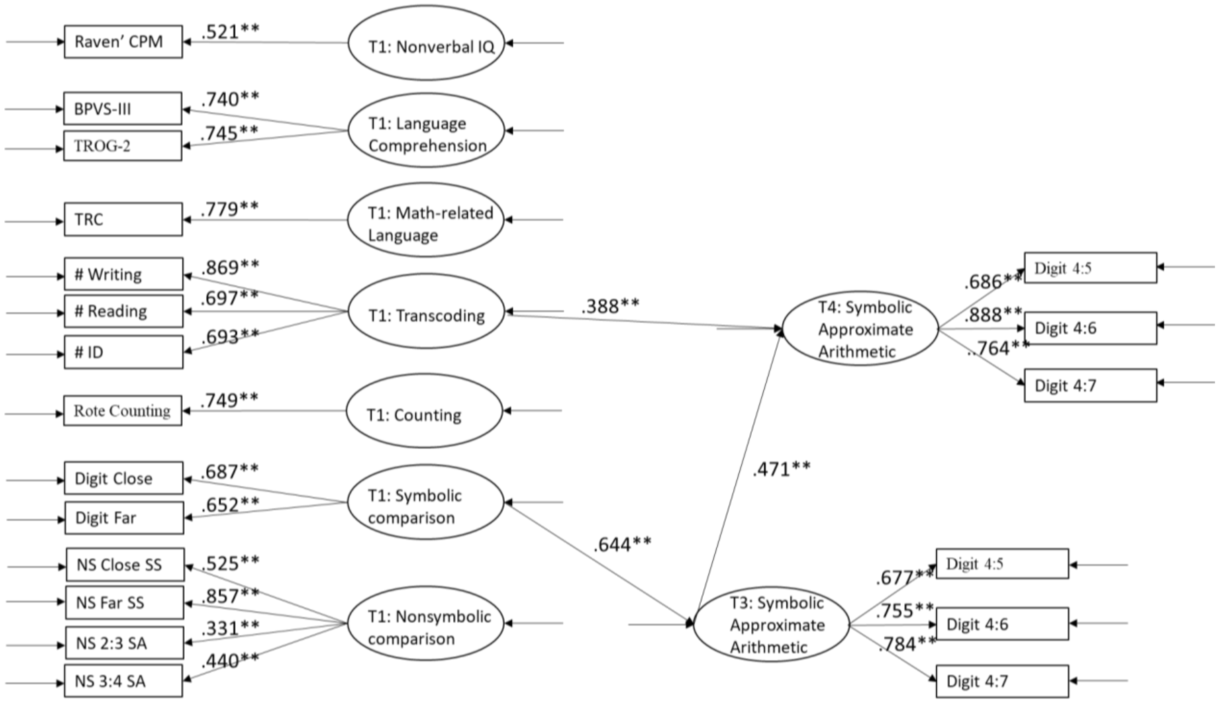


Figure 7.7. Prediction of symbolic approximate arithmetic at Time 4 by Time 1 base model and Time 3 autoregressor. * p < .05. ** p < .01.

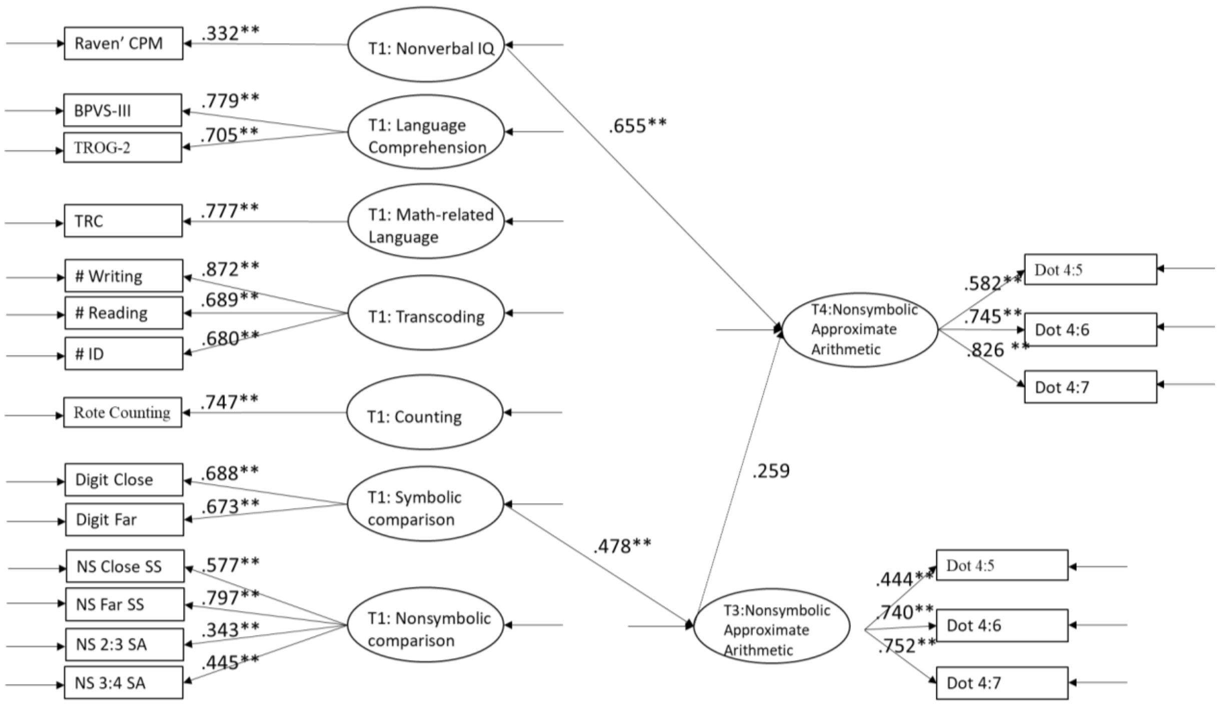


Figure 7.8. Prediction of nonsymbolic approximate arithmetic at Time 4 by Time 1 base model and Time 3 autoregressor. * p < .05. ** p < .01.

At Time 5, the model assessing the prediction of symbolic approximate arithmetic provided an adequate fit to the data, χ^2 (149) = 178.178, *p* = .052, *RMSEA* = .036 (90% CI = .000 - .055), *CFI* = .949, *SRMR* = .078. Figure 7.9 shows that symbolic magnitude comparison at Time 1 was the only unique predictor of the autoregressor symbolic approximate arithmetic at Time 3, and the autoregressor and transcoding at Time 1 were the two unique predictors of symbolic approximate arithmetic performance at Time 5 (42.3% of variance of symbolic approximate arithmetic at Time 3 and 54.9% of variance of arithmetic at Time 5 explained).

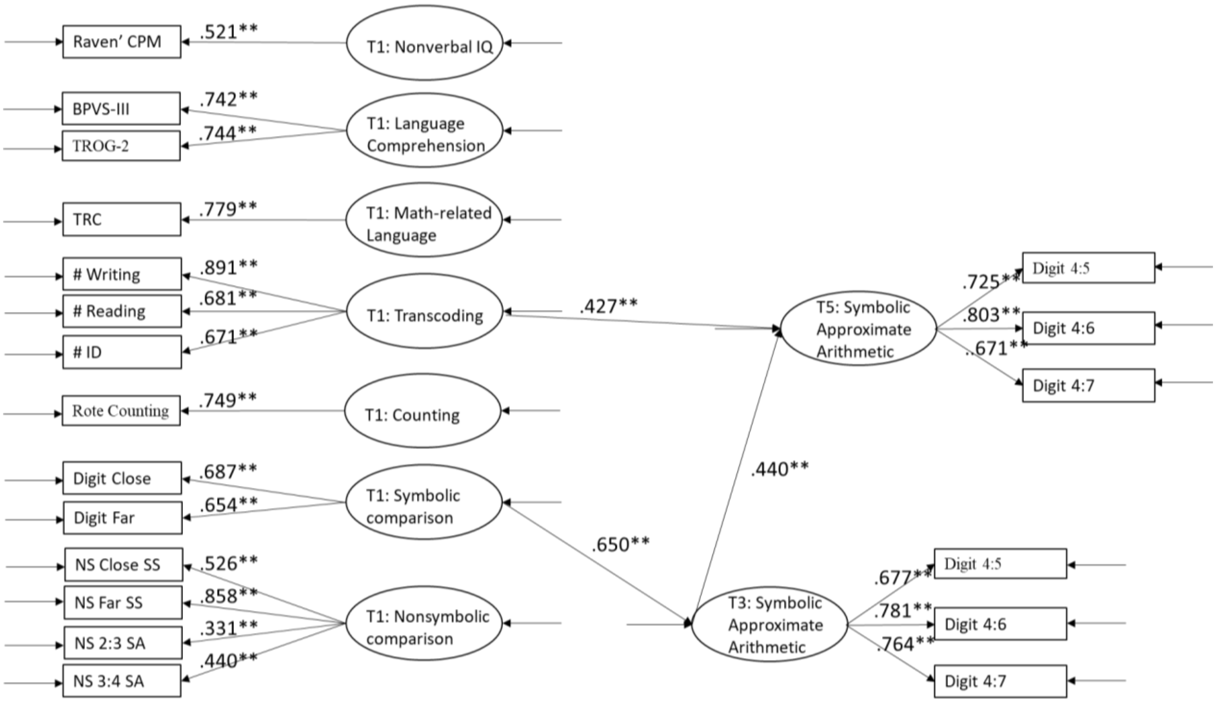


Figure 7.9. Prediction of symbolic approximate arithmetic at Time 5 by Time 1 base model and Time 3 autoregressor. * p < .05. ** p < .01.

Likewise, nonsymbolic approximate arithmetic at Time 5 (Figure 7.10) was predicted by the autoregressor Time 3 and transcoding at Time 1 which was the stronger predictor. The model fit was excellent, χ^2 (149) = 160.343, *p* = .248, *RMSEA* = .023 (90% CI = .000 - .046), *CFI* = .976, *SRMR* = .080. The autoregressor was predicted by symbolic magnitude comparison assessed at Time 1. Symbolic magnitude comparison at Time 1 explained 28.9% of variance in nonsymbolic approximate arithmetic at Time 3 and the autoregressor at Time 3 and transcoding at Time 1 explained 40.0% of variance of nonsymbolic approximate arithmetic at Time 5.

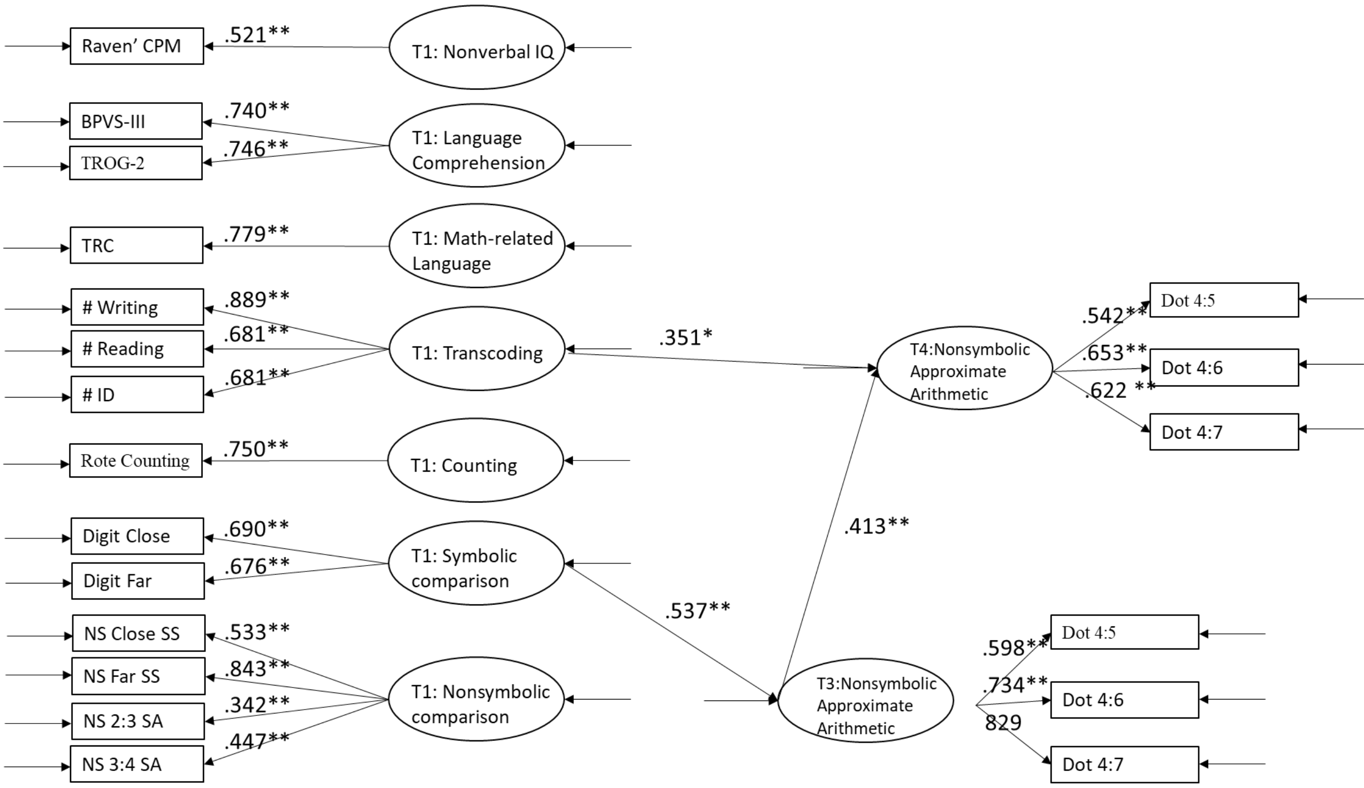


Figure 7.10. Prediction of nonsymbolic approximate arithmetic at Time 5 by Time 1 base model and Time 3 autoregressor. * p < .05. ** p < .01.

7.3 Conclusion.

The main focus of the chapter was to investigate the approximate arithmetic as used in Gilmore et al. (2007). In general, the results of these analyses replicated the findings that young children can perform nonsymbolic as well as symbolic approximate arithmetic with accuracy above chance, and that nonsymbolic approximate arithmetic was easier for young children than symbolic approximate arithmetic. Gilmore and colleagues (2007) sampled a relatively wide age range (five to six year old children) and reported accuracy levels of 70% or more for symbolic approximate arithmetic. In the current study, children at three time points were assessed from five years, six months to six years, four months. The results showed that the younger children at were not as accurately as reported in Gilmore et al. (2007) and only children at Times 4 and 5 (six years and over) showed similar performance as in the Gilmore study.

Second, the results revealed ratio effects (4:7 > 4:6 > 4:5) for symbolic and nonsymbolic approximate arithmetic across all three time points. Children performed more accurately on ratios with a large difference (4:7) than ratios with a small difference (4:5) similar to Gilmore et al. (2007). The results suggest an increase in the ratio effects over time, as indicated by the rise of the effect size partial eta-squared.

Additionally, the chapter examined the correlation between measures of approximate arithmetic and the correlation between these and measures of exact arithmetic. First, symbolic and nonsymbolic approximate arithmetic were significantly related (r_{T3} = .557, r_{T4} = .492 and r_{T5} = .520, *p's* < .001). Composite scores of symbolic and nonsymbolic approximate arithmetic were examined in relation to composite scores of the TOBANS subtasks. Both symbolic and nonsymbolic approximate arithmetic correlated significantly with TOBANS, the correlations with symbolic approximate arithmetic were higher than with nonsymbolic. It seems that symbolic approximate arithmetic is more closely related to traditional arithmetic tasks. Children may not only rely on mental arithmetic when performing nonsymbolic approximate arithmetic but performance may also be underpinned by broader magnitude estimation processes which are not necessary when solving conventional, exact arithmetic. It must be critically mentioned, that nonsymbolic approximate arithmetic was only weakly related to exact arithmetic and symbolic approximate arithmetic moderately.

Questions remain regarding the relationship between approximate and exact arithmetic. To what extent do the correlations noted above suggest common processes in the performance of approximate and exact arithmetic? Might the moderate correlation between symbolic approximate arithmetic and exact arithmetic tasks be explained by common demand on symbol identification? Note in general that these are zero-order correlations which allow only limited interpretation.

In regards to the structure and relationship of the symbolic and nonsymbolic approximate arithmetic tasks, no shift from a two-factor model towards a single-factor model, as found in the development of magnitude comparison, could be noted. At Time 3, only the two-factor model adequately fitted the data suggesting that symbolic and nonsymbolic approximate arithmetic tasks loaded on separate factors at five years and six months. Likewise, it seems that approximate arithmetic at Time 4 also comprises of two separate constructs: symbolic and nonsymbolic approximate arithmetic. The same result was found at Time 5 (age six years, four months) though the model fit of the unitary factor model was improving over time. The findings indicate that the current study may have failed to measure the sensitive period for this switch from the two-factor towards the one-factor model. It may be possible that the shift happens later, perhaps in the second year of formal schooling. It could be that, similar to the development of magnitude comparison, children have to master their knowledge of simple arithmetic problems. As demonstrated in Chapter 3, the shift in magnitude comparison did not occur until children had basic mastery of the Arabic numeral system suggesting that pre-school children's performance on symbolic and nonsymbolic magnitude comparison may rely on distinct forms of representation. Arguably, children have to master simple arithmetic skills before a shift in approximate arithmetic towards the unitary factor may happen.

Alternatively, symbolic and nonsymbolic approximate arithmetic may never form one factor due to the nature of the tasks being so distinct. Symbolic approximate arithmetic may provide underpinning for the more traditional exact arithmetic route whereas nonsymbolic may depend more on magnitude estimation processes. Further research is needed to assess if and when this shift may occur.

Last but not least, the chapter also examined to what extent approximate arithmetic may be predicted by the Time 1 base model. At Time 3, both symbolic and nonsymbolic approximate arithmetic were uniquely predicted by Time 1 symbolic magnitude comparison. At Time 4, symbolic approximate arithmetic was uniquely predicted by the autoregressor (Time 3) and by transcoding (Time 1) whereas nonsymbolic approximate arithmetic at Time 4 was predicted by the autoregressor Time 3 and nonverbal intelligence (Time 1, the stronger predictor). Symbolic approximate arithmetic assessed at Time 5 was predicted by the autoregressor at Time 3 and transcoding at Time 1. Likewise, children's Time 5 nonsymbolic approximate arithmetic performance was predicted by the autoregressor at Time 3 and transcoding (Time 1).

Some tentative conclusions may be drawn. It seems that symbolic magnitude comparison is crucial for children's understanding of early approximate arithmetic, whether symbolic or nonsymbolic. Why would that be? Easier to interpret is the early and specific involvement of nonverbal ability (nonverbal reasoning) in later nonsymbolic approximate arithmetic. However, over time we observe this involvement to be overtaken by transcoding, which operates as a consistent longitudinal predictor of both symbolic and nonsymbolic approximate arithmetic in six year olds. It would be valuable to research this relationship in detail assessing to what extent the knowledge of the symbol system underpins arithmetic generally, and to what extent the relationships observed here are due to the shared cognitive processes of magnitude comparison. Also to be considered is the cumulative effect of increasing practice in exact arithmetic in the early years of formal schooling.

Overall, the symbolic approximate arithmetic task introduced by Gilmore et al. (2007) has proven useful when testing pre-school children who have an immature knowledge of exact arithmetic and had no formal training of exact arithmetic, yet are able to perform above chance level on approximate arithmetic based on double digit numbers. Current findings indicate that general magnitude estimation processes underlie this ability. However, detailed analyses do not support the statement made by Gilmore et al. (2007) that children 'used nonsymbolic number representations to solve symbolic problems' (p.590). There is no direct evidence to support this statement.

Confirmatory factor analyses of current data show that a two-factor model, in which symbolic and nonsymbolic approximate arithmetic are identified as separate latent variables, is preferred across the testing period observed in both studies. Longitudinal analyses showed a common dependence on symbolic comparison and a specific contribution of nonverbal ability to nonsymbolic approximate arithmetic, which is then superceded by a common influence of early symbol transcoding.

Taken together, these findings indicate that performance on ANS comparison tasks, as well as symbolic transcoding, should be taken into account when examining the relationship between approximate arithmetic and exact arithmetic. It will be interesting for future research to explore the development of symbolic and nonsymbolic approximate arithmetic at later stages, and to see whether there will be a shift, similar to magnitude comparison, from a two-factor model with symbolic and nonsymbolic approximate arithmetic as distinct constructs towards a unitary model with one general approximate arithmetic construct.

Chapter 8. The Importance of Children's Number Estimation on Arithmetic

There is evidence that there is a shift from a logarithmic to a linear distribution of numerical magnitude representations in children between five and eight years (Booth and Siegler, 2008; Siegler et al., 2009), and that older children perform better on number lines (as indicated by the difference between actual position and children's estimated position) than younger children, and that performance on the number line is significantly associated with arithmetic skills (Siegler and Booth, 2004; Booth and Siegler, 2006, 2008). However, these findings come mostly from cross-sectional studies.

Some studies identified the involvement of language, counting and magnitude estimation as developmental associates of early arithmetic (Praet et al., 2013, Wiese, 2003). Indeed, Praet and Desoete (2013) explored the relationship between arithmetic and children's estimation using number words, dots and Arabic numerals, adding language as a covariate (from kindergarten till grade two). The results revealed that Arabic numerals were more linearly distributed than number words and that language explained kindergartener's arithmetic performance, but not the growth of arithmetic. Children's untimed math performance was predicted by number line estimation. There is further evidence for the importance of number estimation in the development of arithmetic skills (Muldoon et al., 2013). They authors noted that five-year-olds' counting ability was the largest contributor to children's math performance and only linear fit of number estimation on the 0-20 scale at 5 years and linear fit of number estimation on the 0-100 scale at six years made a significant contribution. Some studies propose that, rather being a predictor of mathematical achievement, number line acuity and math performance both influence each other during development from pre-school through early school years (Friso-van den Bos et al., 2014; LeFevre et al., 2013).

There is some evidence that children know and rely on multiple numerical representations (Siegler and Opfer, 2003). The authors further showed that children rely more on linear representations rather than intuitive, logarithmic ones and that numerical context (0-100 or 0-1000 scale) affects the type of representation. The study examined performance on numerical estimation tasks in second, fourth and sixth grade children and adults.

The present study examined the nature and development of young children's numerical estimation abilities and its relation to arithmetic over the time course from pre-school to the conclusion of the first year of formal schooling. The relation between numerical estimation, counting and arithmetic proposed by Muldoon et al. (2013) were tested, and the predictive value of numerical estimation within a comprehensive model of the development of arithmetic was further explored.

8. 1 Methods.

8.1.1 Participants.

The same participants were used as described in Chapter 2 (p. 42)

8.1.2. Materials.

Children were assessed on the following measures.

8.1.2.1 Baseline model taken at Time 1.

The following tasks (independent variables) were administered individually to the four nursery classes in the summer term of the nursery age (4 years; see Chapter 4, pp. 76-79 for more details): Nonverbal intelligence (Raven's CPM; Raven et al., (1993), grammatical ability (TROG-2; Bishop, 2003), vocabulary (BPVS - III; Dunn et al., 2010), specific math-related language ability (TRC), transcoding (Number Identification, Number Writing and Reading Arabic numerals), rote counting, magnitude comparison and arithmetic (simple addition problems).

8.1.2.2 Numerical Estimation.

Time 1. Children's numerical estimation skill was assessed using the Number-to-Position task of the traditional number line task (Siegler and Opfer, 2003; Whyte and Bull, 2008). Nine 20 cm lines with a start anchor point of 0 (left end of line) and end anchor point of 10 (right end) were presented to the child. The child was asked to mark the position of the target number on the number line. The target numbers (1, 2, 3, 4, 6, 7, 8, or 9) were presented verbally in random order. Furthermore, children were also assessed on two items (13 and 16) on the number line scale 0-to-20.

The experimenter explained the number line and start and end anchor beforehand. One practice trial (5) was given to familiarise the child with the task.

Feedback on how to mark the position was given to the child due to the fact that many children seemed not to fully understand the task.

Time 3. Testing procedure was the same as in Time 1 but the difficulty level of the task was adjusted for age. Two number line ranges were included (0-to-10 and 0-to-20). Five 25 cm lines for each number range were presented with the appropriate start and end anchor points. The target numbers for 0-to-10 were 6, 2, 9, 12 and 7 and for 0-to-20 were 3, 8, 14, 6 and 17. Five practice trials (number 3 and 5 for 0-10, 12 for 0-20 and 72 and 29 for 0-100) were presented prior to test trials.

Time 5. Testing procedure was the same as in Time 3. The number line ranges were 0-to-10 and 0-to-20. Each number line was 20 cm in length and was presented with the appropriate start and end anchor points. The target numbers for 0-to-10 were 6, 2, 9, 1, 7, 3, 8 and 4 and for 0-to-20 were 3, 8, 14, 6, 17, 4 and 12. Four practice trials (number 5 for 0-10 and numbers 9, 15 and 10 for 0-20) were tested prior to test trials.

8.1.2.3 Arithmetic measures taken at Time 5.

Arithmetic. *Fluency.* Children's speeded arithmetic skills (fluency) was assessed using the 'addition', 'addition with carry' and 'subtraction' subtask' subtests of the TOBANS (Brigstocke et al., 2016). Children were asked to complete as many arithmetic problems as possible in one minute. In the 'addition' subtask, children were presented with simple addition problems with sums less than ten and in the 'addition with carry' subtask the sums were bigger than ten but smaller than twenty. One point was awarded even if the numeral was written backwards (maximum score $_{\text{addition, subtraction}}$ = 90; maximum score $_{\text{addition with carry}}$ = 30). This task was administered as a group task.

Accuracy. Children's basic arithmetic skills were assessed using the Numerical Operations subtest of the second edition of the Wechsler Individual Achievement Test (WIAT-II; Wechsler, 2005). The first six items (identifying and writing Arabic numerals) were in order to provide a more specific and conventional measure of arithmetic. The test was executed according to the manual and children were allowed to complete the task in their own time (maximum score = 25).

8.1.3 Procedure.

Testing was carried out three times over a 25-month period from the summer term of nursery through to the summer term of Year One. Wherever possible, each child was seen by the same experimenter. The author was assisted by several research assistants from undergraduate psychology classes. They were trained on how to administer the test battery and were given instructions on how to work with young children. Children were tested individually at Times 1 and 2, and at Times 3, 4 and 5 individually or in small groups in a separate room or another quiet place in the school. Numerical estimation was individually assessed at all three time points. Each child met with the experimenter ideally two to four days in a row, depending on the number of blocks, to avoid lack of motivation or concentration. If testing in groups, the ratio of experimenters to children was 1:3.

Preliminary to the testing, the experimenters attended at least one day in each class so that the children got to know them and felt more comfortable around them. Moreover, the experimenters told each child that they would play games and asked questions such as “How are you?” or “How old are you?”.

All unstandardized tests included practice items. Concerning feedback, children received only concrete feedback on their performance for practice items and general praise and encouragement throughout the tests.

8.2 Results.

Descriptive statistics for the number estimation task are shown in Table 8.1. To answer the research question regarding the relationship between number estimation and early arithmetic performance, descriptive statistics are first examined, followed by repeated-measures ANOVAs run on IBM SPSS Statistics 22. The dependent variable was children’s performance on arithmetic tasks at each time point and the independent variable consisted of children’s linear and logarithmic fit to number line estimates (difference between child’s mark and target position); factors included time and scale of number line. Then, hierarchical regressions established the relationship between early arithmetic, linear and logarithmic fit to number line estimates (difference between child’s mark and position of target number) and counting. Further analyses examined the extent to which number estimation may predict early arithmetic skills, based on SEM path models using MPlus Version 7. In the SEM path models, the independent variables comprised the Time 1 baseline model, expanded to include children's number estimation (difference between child's mark and position of target number) was regressed onto the dependent latent factor of children's arithmetic scores at Time 5 (TOBANS and WIAT subtasks). To visually simplify the path models, the coefficients of the relations between the factors are not presented but can be found in Appendix 25.

Estimation data was analysed for changes in both linearity and accuracy (error). Linear and logarithmic functions (using difference scores between child's mark and target position) were fitted using the equations employed by Muldoon et al. (2013): $y = slopex + b$ (linear function) and $y = clnx + b$ (logarithmic function). Error denotes error percentage (percentage of how much the child's answer deviates from the target position) and absolute error (difference between child's answer and target number). Error percentage was calculated using the equation *error = (child's estimate – number to be estimated) / number line scale x 100*. Children's absolute error values were used in the SEM path model analysis.

8.2.1 Descriptive Statistics

Descriptive statistics are shown in Table 8.1. At Time 1, children's performance on 0-10 scaled items was most accurate for number '6' (19.67% error) and number '8' (42.34% error) showed the worst performance. A different pattern was found for Time 3 with number '3' and number '1' (14.20% error and 8.88% error respectively) being the most accurate item and performance on number '6' at Times 3 and 5 being the least accurate (33.14% error and 27.69% error respectively). Children's accuracy on 0-10 items increased over time. Interestingly, it seems that children were more accurate on small numbers compared to larger numbers. The error percentage at Time 1 for items 1-7 was overall smaller than items '8' and '9'. Surprisingly, error seems to be relatively higher than expected for numbers '1' and '2' at Time 1. At Times 3 and 5, it seems that the breakpoint may be around number '5' with numbers in the subitizing range (1-4) being easier to estimate than large numbers reflecting most likely the frequency input (Dehaene and Mehler, 1992).

In regards to performance on 0-20 scaled items, children performed most accurately on number '3' (11.29% error Time 3 and 7.37% error at Time 5) and performance on number '17' was the least accurate (28.99% error at Time 3 and

21.16% error at Time 5). Children’s accuracy on 0-20 items improved over time and children’s estimates on the 0-20 scale were more accurate than on the 0-10 scale. Children‘s estimates on numbers smaller than ten was more accurate on the 0-20 scale compared to the 0-10 scale.

Table 8.1

Mean and standard deviations of predictor and criterion measures from all testing sessions

		Time 1	Time 3	Time 5
		M (SD)	*M (SD)*	*M (SD)*
Nonverbal IQ	Raven's CPM	6.45 (1.57)		
Language Comprehension	TROG-2	3.15 (2.63)*		
Vocabulary	BPVS-III	58.26 (16.77)*		
Math-related Language	TRC - more	5.90 (2.00)*		
Numerical Knowledge	Number Writing	6.86 (5.79)*		
	Number Reading	8.02 (2.67) [45]*		
	Number Identification	7.26 (2.60)*		
	Rote Counting	14.78 (12.84)*	1 to 40: 34.84 (8.51) [60] 94 to 110: 10.30 (5.34) [26] 25 back.: 7.01 (7.73) [12]*	
Magnitude Comparison	Digit Close	10.11 (3.21) [7]		
	Digit Far	10.73 (3.59) [7]		
	NS FS Close	10.26 (2.16) [1]		
	NS FS Far	12.87 (2.57) [17]		

	NS FS 3:4	10.76 (2.72) [4]		
	NS FS 5:6	10.58 (2.29) [3]		
	NS SA Close	10.29 (2.28) [1]		
	NS SA Far	13.24 (2.43) [23]		
	NS SA 2:3	11.42 (2.42) [5]		
	NS SA 3:4	10.81 (2.03) [2]*		
Arithmetic	Addition Tasks	6.33 (3.21) [17]*	A: 5.23 (2.51) [3]	
			B: 5.15 (2.55) [8]*	
	TOBANS			
	Addition			12.74 (8.66)
	Addition w/ carry			5.07 (5.01) [1]
	Subtraction			8.44 (5.10)
	WIAT			4.00 (2.27)

		Time 1		Time 3		Time 5	
		Position	Error %	Position	Error %	Position	Error %
		M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Number Estimation	0-10						
	1	3.2 (2.7)	25.92 (23.0)	1.1 (2.4)	14.20 (18.7)	.4 (1.0)	8.88 (7.9)
	2	3.9 (2.7)	26.13 (19.2)	1.3 (2.1)	17.41 (14.4)	.9 (1.1)	13.83 (7.1)
	3	4.1 (2.7)	24.09 (16.3)			1.2 (.8)	19.18 (5.9)
	4	4.1 (3.0)	22.40 (19.4)			1.8 (1.3)	24.21 (7.5)
	6	10.8 (2.3)	19.67 (16.4)	3.0 (2.1)	33.14 (15.3)	3.3 (1.4)	27.69 (12.8)
	7	5.2 (2.4)	23.53 (18.4)	4.6 (2.9)	31.94 (20.1)	4.6 (1.8)	25.01 (17.1)

8	10.0 (4.9)	42.34 (31.9)			5.8 (2.0)	22.01 (20.1)
9	5.9 (2.6)	32.67 (23.9)	6.3 (3.6)	32.09 (31.1)	6.4 (2.3)	25.83 (23.2)
0-20						
3			3.0 (3.8)	11.29 (15.3)	2.0 (1.6)	7.37 (6.1)
4					2.8 (1.8)	8.82 (6.3)
6			5.7 (4.3)	15.4 (14.8)	4.6 (2.1)	10.68 (6.6)
7						
8			8.5 (5.5)	22.29 (16.3)	8.0 (3.4)	13.70 (9.9)
12					9.8 (3.2)	15.30 (11.7)
13	10.5 (5.4)	25.25 (15.3)				
14			11.4 (6.0)	25.82 (19.7)	11.2 (3.0)	16.56 (12.1)
16	10.8 (5.3)	30.82 (20.4)				
17			12.0 (5.6)	28.98 (23.9)	12.8 (2.8)	21.16 (14.0)

Notes. M = mean age. SD = standard deviation * *individually administered tasks. The number of children scoring at maximum level are shown in square brackets. All scores are presented as raw scores. For the Magnitude Comparison Tasks: NS = nonsymbolic. FS = fixed size trials. SA = surface-area matched trials.*

8.2.2 The associations between children's number line estimation and early arithmetic.

Descriptive statistics of children's overall error percentage of estimates as well as linear and logarithmic function fits to number estimation line using absolute errors (difference between child's mark and target position) are shown in Table 8.2.

The analysis of children's estimates at Time 1 showed a significantly better linear fit (R^2_{lin}) than logarithmic fit (R^2_{log}), $t(99) = 7.168$, $p < .001$, $r = .980$. Next, a 2 by 2 repeated-measures ANOVA with Time (Times 3 and 5) and Scale (0-10 and 0-20) as factors and the linear fit (R^2_{lin}) as dependent variable revealed that linear fit significantly improved over time ($F(1,106) = 30.604$, $p < .001$, $\eta_p^2 = .224$; T3: $M = .650$, $SD = .02$; T5: $M = .813$, $SD = .01$), but the linear performance on scale 0-10 ($M = .752$, $SD = .02$) and 0-20 ($M = .710$, $SD = .01$) did not differ ($F(1,106) = 2.705$, $p = .103$, $\eta_p^2 = .025$). There was also a significant interaction between time and scale ($F(1,106) = 4.407$, $p = .038$, $\eta_p^2 = .040$). Post-hoc t-tests showed that at Time 3, performance on 0-20 ($M = .607$, $SD = .03$) was less linear than 0-10 ($M = .693$, $SD = .03$), $t(113) = 2.501$, $p = .014$, whilst there was no difference on linearity at Time 5, $t(114) = -.244$, $p = .807$, 0-10: $M = .812$, $SD = .02$, 0-20: $M = .813$, $SD = .02$).

A further 2 by 2 repeated-measures ANOVA investigated the fit of the logarithmic function (R^2_{log}), with Time (Times 3 and 5) and Scale (0-10 and 0-20) as the factors. The results showed that the logarithmic fit improved over time ($F(1,106) = 22.333$, $p < .001$, $\eta_p^2 = .174$; T3: $M = .627$, $SD = .02$; T5: $M = .757$, $SD = .01$). There was also a significant main effect for scale ($F(1,106) = 7.603$, $p = .007$, $\eta_p^2 = .067$), with a more logarithmic performance on the 0-20 scale ($M = .724$, $SD = .02$) compared to 0-10 ($M = .660$, $SD = .01$). Again, the interaction was significant ($F(1,106) = 10.512$, $p = .002$, $\eta_p^2 = .090$). Post-hoc t-tests showed that there was no effect of scale at Time 3 ($t(113) = -.012$, $p = .991$, 0-10: $M = .626$, $SD = .03$, 0-20: $M = .628$, $SD = .03$), but at Time 5 ($t(114) = -.5.508$, $p < .001$), there was a less strongly logarithmic performance on 0-10 ($M = .690$, $SD = .02$) scale than 0-20 ($M = .820$, $SD = .02$) scale.

Table 8.2

Estimation error, R^2_{lin} and R^2_{log} for 0-10 and 0-20 scale of number estimation task at all three time points

	0-10			0-20		
Measure	T1	T3	T5	T1	T3	T5
	Analysis of individual children's estimates (mean values)					
Error (%)	27.29	32.68	20.75	28.03	40.60	13.45
R^2_{lin}	.329	.679	.815		.593	.822
R^2_{log}	.284	.612	.697		.613	.827
	Analysis by group (median values)					
R^2_{lin}	.476	.867	.935		.967	.985
R^2_{log}	.414	.707	.758		.978	.975

Notes. Although children were assessed on two 0-20 scaled targets at Time 1, only errors could be analysed.

8.2.3 The quality of children's number line estimation as a predictor of early arithmetic.

To answer the question of whether these changes in children's estimates are related to changes in children's performance on early arithmetic, correlations between arithmetic scores (composite scores of the arithmetic subtasks assessed at each time point) at all three time points and function fits of number estimates for the scales 0-10 and 0-20 were compared for both linear and logarithmic function. As shown in Table 8.3, neither the linear nor the logarithmic estimates of the scale 0-10 were significantly related to arithmetic at Time 1. Significant correlations were observed between arithmetic scores and R^2_{lin} values on the 0-20 scale and a weak correlation on the 0-10 scale at Time 3, whereas neither 0-10 values nor 0-20 values were significantly correlated to arithmetic at Time 5. A similar pattern emerged for logarithmic fits. R^2_{log} values on the 0-20 scale were significantly related with arithmetic at Time 3. At Time 5, there were no significant correlations with neither 0-10 nor 0-20.

Table 8.3

Correlations between the linear and logarithmic function fits and arithmetic at Times 1, 3 and 5

	Time 1			Time 3			Time 5		
	1	2	3	1	2	3	1	2	3
1. Linear fit 0-10	---	----	.128	---	.283**	.210*	---	-.003	.098
2. Linear fit 0-20		----	----		----	.417**		----	.106
3. Arithmetic			----			----			----
1. Logarithmic fit 0-10	---	----	.122	---	.302**	.163	---	.080	.060
2. Logarithmic fit 0-20		----	----		----	.352**		----	.133
3. Arithmetic			----			----			----

Notes. Pearson product-moment correlation coefficient. Arithmetic variables entered are composite scores of raw scores. * $p < .05$. ** $p < .01$

Next, it was investigated to what extent the relationship between children's number line estimation skills (linear and logarithmic function fit; independent variable) and early arithmetic (dependent variable; composite scores of arithmetic subtasks) was mediated by other covariates including counting ability and transcoding (independent variables). The analysis focused on Times 1 and 3 performance. At Time 1, most children (85%) could count up to ten, but only 18% could count up to twenty. At Time 3, 96.5% of the children could count up to 10, 86% up to twenty and 52.6% could count up to 40. The correlation between arithmetic and rote counting at Time 1 was significant (r = .28), as were the correlations between arithmetic at Time 3 and counting to 40 (*r* = .34), counting from 94 to 110 (*r* = .53) and counting backwards from 25 (*r* = .65).

Regression analyses were used to examine to what extent the relationship between number line estimation and early arithmetic was mediated by counting skills and transcoding at Time 1. The Time 1 correlation matrix for counting, transcoding tasks and linear and logarithmic fit for the 0-10 scale is shown in Table 8.4.

Table 8.4

Correlations between the linear and logarithmic function fits, counting and transcoding at Time 1

	1.	2.	3.	4.	5.	6.
1. Counting	----	.341**	.228*	.475**	.124	.123
2. # ID		----	.562**	.586**	.206*	.181
3. # Reading			----	.579**	.254*	.243*
4. # Write				----	.261**	.254*
5. Linear fit 0-10					----	.980**
6. Logarithmic fit 0-10						----

Notes. Pearson product-moment correlation coefficient. All variables entered were of raw scores.
** p < .05. ** p < .01*

Hierarchical regressions were conducted in which counting was entered in step one, transcoding tasks (number reading, writing and identification scores were entered seperately) were added in the second step and R^2 values in step three (separate regressions for linear and logarithmic function fits). As shown in Table 8.5, counting was significantly predicting arithmetic (*ß* = .281, *t*(98) = 2.903, *p* = .005, r^2 = .281). Adding transcoding, was significantly improving the model, with number identification being the only unique predictor (*ß* = .514, *t*(97) = 5.387, *p* < .001, r^2 = .480, r_{change}^2 = .212, F_{change}(1, 97) = 29.023, *p* < .001).

Table 8.5

Hierarchical Regressions for Counting, Transcoding and linear and logarithmic function fits and Arithmetic at Time 1

	Linear Fit			Logarithmic Fit		
	B	*SE B*	*ß*	*B*	*SE B*	*ß*
Step 1						
Constant	5.291	.47		5.291	.47	
Counting	.07	.02	.28**	.07	.02	.28**
Step 2						
Constant	2.703	1.03		2.703	1.03	
Counting	.013	.03	.05	.013	.03	.05
Transcoding						
# Id	.208	.07	.38**	.208	.07	.38**
# Reading	.045	.14	.04	.045	.14	.04
# Write	.211	.14	.18	.211	.14	.18
Step 3						
Constant	2.750	1.04		2.754	1.04	
Counting	.013	.03	.05	.013	.03	.05
Transcoding						
# Id	.210	.07	.38**	.210	.07	.38**
# Reading	.046	.14	.04	.044	.14	.04
# Write	.216	.14	.18	.216	.14	.18
0-10	-.364	1.05	-.03	-.405	1.12	-.03

*Notes. Linear Fit: Step 1 R^2 = .079, p = .005, Step 2 ΔR^2 = .219, p < .001, Step 3 ΔR^2 = .001, p = .729. Logarithmic Fit: Step 1 R^2 = .079, p = .005, Step 2 ΔR^2 = .219, p < .001, Step 3 ΔR^2 = .001, p = .734. * p < .05. ** p < .01.*

The R^2_{lin} values on the 0-10 scale did not significantly contribute to the variance on arithmetic scores ($ß$ = -.028, $t(96) = -.317$, $p = .752$, $r^2 = -.032$), r_{change}^2 = .001, $F_{change}(1, 96) = .101$, $p = .752$).

A second hierarchical regression model was conducted adding R^2_{log} values on the 0-10 scale in step three (shown in Table 8.5). The same pattern was observed that counting was a significant predictor of arithmetic at Time 1, adding measures of transcoding, and number identification in particular, in step two significantly improved the model, but adding R^2_{log} values on the 0-10 scale did not significantly improve the model ($ß$ = -.026, $t(96) = -.289$, $p = .773$, $r^2 = -.029$), $r_{change}{}^2 = .001$, $F_{change}(1, 96) = .084$, $p = .773$).

The Time 3 correlation matrix for counting (counting to 40, from 94-110 and backwards from 25 were entered separately), transcoding tasks (entered separately) and linear and logarithmic fit for the 0-10 scale are shown in Table 8.6. At Time 3, a set of hierarchical regressions were run in which counting (counting to 40, counting from 94 to 110 and counting backwards from 25) was entered in step one, the transcoding tasks number reading, writing and identification were added seperately in the second step and R^2 values on the 0-10 and 0-20 scale were added in step three (separate regressions for linear and logarithmic function fits). The linear hierarchical regression (Table 8.7) showed that both counting from 94 to 110 ($ß$ = .269, $t(106) = 3.242$, $p = .002$, $r^2 = .300$) and counting backwards from 25 ($ß$ = .508, $t(106) = 6.344$, $p < .001$, $r^2 = .525$) were significantly predicting arithmetic, but counting to 40 was not significant ($ß$ = .040, $t(106) = .500$, $p = .618$, $r^2 = .049$). The addition of transcoding made a significant improvement on the model ($ß$ = .387, $t(105) = 4.439$, $p < .001$, $r^2 = .397$, $r_{change}{}^2 = .082$, $F_{change}(1, 105) = 19.7011$, $p < .001$). Neither R^2_{lin} values on the 0-10 scale ($ß$ = .058, $t(103) = .836$, $p = .405$, $r^2 = .082$) nor R^2_{lin} values on the 0-20 scale ($ß$ = .068, $t(103) = .913$, $p = .363$, $r^2 = .090$) were significantly predicting arithmetic ($r_{change}{}^2 = .009$, $F_{change}(1, 103) = 1.089$, $p = .340$).

The Time 3 correlation matrix for counting, transcoding tasks and linear and logarithmic fit for the 0-10 scale are shown in Table 8.6. At Time 3, a set of hierarchical regressions were run in which counting (counting to 40, counting from 94 to 110 and counting backwards from 25) was entered in step one, the transcoding tasks number reading, writing and identification were added in the second step and R^2 values on the 0-10 and 0-20 scale were added in step three (separate regressions for linear and logarithmic function fits).

Table 8.6

Correlations between the linear and logarithmic function fits, counting and transcoding at Time 3

	1	2	3	4	5	6	7	8	9	10
1. Counting to 40	----	.426**	.348**	.405**	.440**	.459**	.202*	.357**	.124	.333**
2. Counting from 94 to 110		----	.461**	.392**	.500**	.417**	.090	.371**	.028	.329**
3. Counting backwards			----	.460**	.472**	.390**	.157	.385**	.086	.297**
4. # ID				----	.689**	.519**	.234*	.304**	.179	.297**
5. # Reading					----	.597**	.190*	.305**	.146	.276**
6. # Write						----	.195*	.379**	.146	.365**
7. Linear fit 0-10							----	.283**	.949**	.337**
8. Linear Fit 0-20								-----	.235*	.955**
9. Logarithmic fit 0-10									----	.302**
10. Logarithmic fit 0-20										----

Notes. Pearson product-moment correlation coefficient. Arithmetic variables entered are composite scores of raw scores.

* $p < .05$. ** $p < .01$

Table 8.7

Hierarchical Regressions for Counting, Transcoding and linear and logarithmic function fits and arithmetic at Time 3

	Linear Fit			Logarithmic Fit		
	B	*SE B*	*ß*	*B*	*SE B*	*ß*
Step 1						
Constant	1.161	1.40		1.161	1.40	
Counting 0-40	.023	.05	.04	.023	.05	.04
Counting 94-110	.279	.09	.27**	.279	.09	.27**
Counting backwards	.420	.07	.51**	.420	.07	.51**
Step 2						
Constant	-3.163	1.60		-3.163	1.60	
Counting 0-40	-.046	.04	-.08	-.046	.04	-.08
Counting 94-110	.174	.08	.17*	.174	.08	.17*
Counting backwards	.313	.07	.38**	.313	.07	.38**
Transcoding						
# Id	.225	.15	.14	.225	.15	.14
# Reading	.108	.13	.08	.108	.13	.08
# Write	.525	.18	.25**	.525	.18	.25**
Step 3						
Constant	-3.628	1.66		-3.743	.168	
Counting 0-40	-.056	.05	-.10	-.052	.05	-.09
Counting 94-110	.163	.09	.16	.173	.09	.17*
Counting backwards	.297	.07	.37**	.309	.07	.37**
Transcoding						
# Id	.207	.15	.13	.205	.15	.13
# Reading	.117	.13	.09	.112	.13	.09
# Write	.478	.18	.23**	.490	.18	.23**
0-10	.776	1.27	.04	1.199	1.39	.06
0-20	1.619	1.37	.09	.910	1.38	.09

*Notes. Linear Fit: Step 1 $R^2 = .479$, $p < .001$, Step 2 $\Delta R^2 = .095$, $p < .001$, Step 3 $\Delta R^2 = .009$, $p = .336$. Logarithmic Fit: Step 1 $R^2 = .479$, $p < .001$, Step 2 $\Delta R^2 = .095$, $p < .001$, Step 3 $\Delta R^2 = .007$, $p = .440$. * $p < .05$. ** $p < .01$.*

The linear hierarchical regression (Table 8.7) showed that both counting from 94 to 110 ($ß$ = .269, t(106) = 3.242, p = .002, r^2 = .300) and counting backwards from 25 ($ß$ = .508, t(106) = 6.344, p < .001, r^2 = .525) were significantly predicting arithmetic, but counting to 40 was not significant ($ß$ = .040, t(106) = .500, p = .618, r^2 = .049). The addition of transcoding (number writing) made a significant improvement on the model ($ß$ = .387, t(105) = 4.439, p < .001, r^2 = .397, r_{change}^2 = .082, F_{change}(1, 105) = 19.7011, p < .001). Neither R^2_{lin} values on the 0-10 scale ($ß$ = .058, t(103) = .836, p = .405, r^2 = .082) nor R^2_{lin} values on the 0-20 scale ($ß$ = .068, t(103) = .913, p = .363, r^2 = .090) were significantly predicting arithmetic (r_{change}^2 = .009, F_{change}(1, 103) = 1.089, p = .340).

Likewise, counting was not a unique predictor of arithmetic once transcoding (number writing) was taken account of and both R^2_{log} values on the 0-10 scale ($ß$ = .043, t(103) = .622, p = .535, r^2 = .061) and R^2_{log} values on the 0-20 scale ($ß$ = .107, t(103) = 1.416, p = .160, r^2 = .138, r_{change}^2 = .009, F_{change}(1, 103) = 1.089, p = .340) failed to make a significant contribution to the prediction of arithmetic scores.

8.2.4 Longitudinal prediction of arithmetic.

The longitudinal SEM analysis of prediction of arithmetic (dependent variable) at Time 5 included the following independent variables assessed at Time 1: the absolute error (difference between child's estimate and number to be estimated) data from the 0-10 scale items of the number estimation task and the Time 1 baseline model as used in previous chapters. Data from 0-20 scale proved difficult to model. It was decided to measure number line estimation as absolute error rather than error percentage because absolute error not only gives an estimate on the accuracy of the answer, but also includes information on whether children underestimated (negative value) or overestimated the target number (positive value).

Latent predictive variables included transcoding (number writing, reading and identification), counting skills (rote counting), the two magnitude comparison constructs (*symbolic* and *nonsymbolic* comparison) as well as two factors for number estimation (small range and large range) and the autoregressor arithmetic (addition task). The factors nonverbal intelligence (Raven's CPM), general language comprehension (BPVS-III and TROG-2), math-related language (TRC) were excluded based on their failure to show significant prediction in previous models, and on a non-positive definite first-order derivative product matrix which is most likely due to having more parameters than the sample size allows. Counting was assessed by only one indicator and thus it was pre-specified with an error term reflecting the reliability of the variable calculated on the sample.

The 0-10 scale items of number estimation task formed two distinct factors because smaller numbers did not load significantly onto one hypothesised factor: small numbers (0-4, children's subitizing range) and large numbers (numbers '6' and '8'). The numbers '7' and '9' were further removed from the large-numbers- factor because of nonsignificant loadings at the alpha level of .01 ('7' (*p* = .201) and '9'(*p* = .021).

The path model shown in Figure 8.1 provided a weak fit to the data, $\chi 2$ (184) = 239.465, *p* = .004, *RMSEA* = .046 (90% CI = .027 - .061), *CFI* = .928, *SRMR* = .085. Similar to previous longitudinal findings, transcoding was the strongest unique predictor of children's performance on arithmetic tasks (56.8% of variance was explained). Performance of large number estimations was also a significant predictor of arithmetic, in a negative direction, consistent with the expectation that the smaller the difference between answer and target number the better the performance on arithmetic at Time 5.

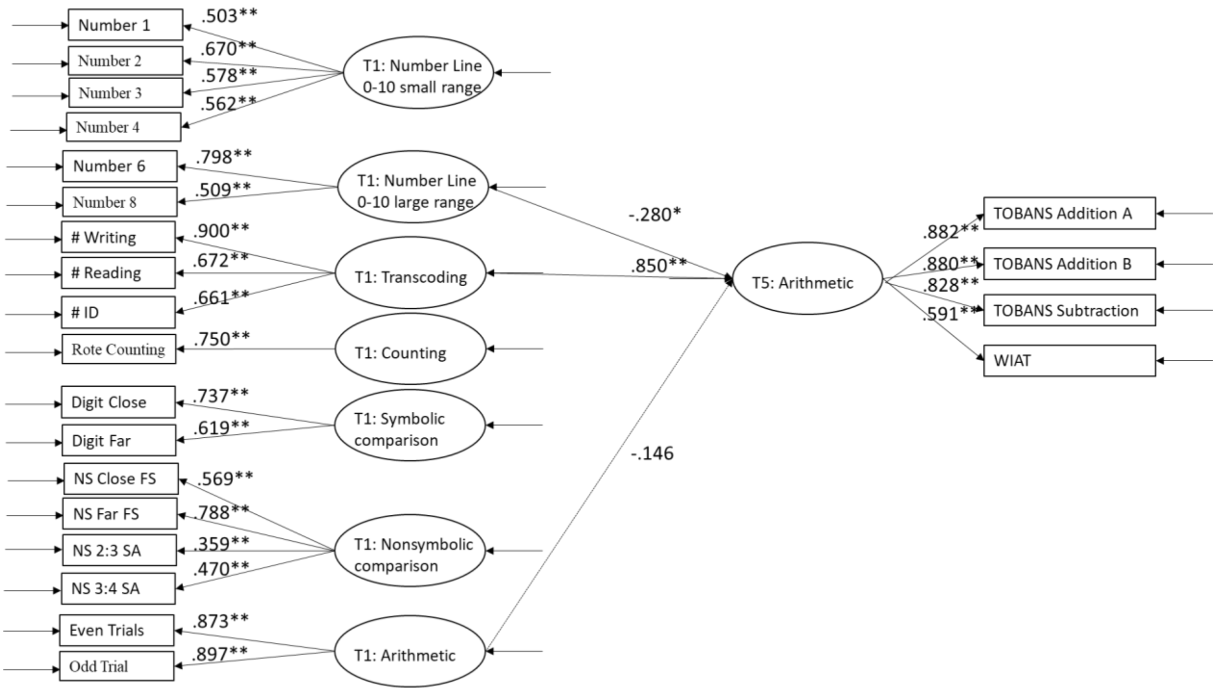


Figure 8.1. Prediction of arithmetic at Time 5 by Time 1 latent factors Transcoding, Counting, Symbolic and Nonsymbolic comparison tasks, number line 0-10 small and number line 0-10 large as well as Time 1 arithmetic autoregressor. * p < .05. ** p < .01.

The fact that the model provided a poor fit and the nonsignificant correlations between latent number line factors with other independent variable factors, and between both number line factors in particular (Table 8.8), limits possible interpretation.

Correlations. The correlation matrix for the latent variables is shown in Table 8.8. As expected, most Time 1 latent factors were significantly correlated with arithmetic at Times 1 and 5, except for both number line factors. Transcoding was the strongest correlate of arithmetic at both time points. Transcoding was further significantly related to counting and symbolic magnitude comparison suggesting that the three numeracy abilities may share some cognitive processes. Both symbolic and nonsymbolic magnitude comparison factors were significantly correlated.

Surprisingly, the number line factor 0-10 for small numbers was only significantly correlated with symbolic magnitude comparison. Children's number line 0-10 scores for large numbers did not significantly correlate with any other factor. Interestingly, the two number line factors were not correlated. The negative correlations, albeit not significant, with the other factors can be explained by the fact that number line was estimated using difference between children's answers and target position. It shows that the bigger the difference between target and answer the poorer the performance on the other factors.

Table 8.8

Correlations between Time 1 baseline model, number estimation 0-10 scale and arithmetic at Time 5 (n = 148)

	1	2	3	4	5	6	7	8
1. Transcoding	---	.679**	.646**	.429**	-.189	.226	.631**	.694**
2. Counting		----	.365*	.342*	-.129	.195	.405**	.463**
3. Symbolic Comparison			---	.531**	-.336*	-.094	.558**	.493**
4. Nonsymbolic Comparison				----	-.167	-.016	.533**	.291**
5. 0-10 small numbers					----	-.095	-.064	-.125
6. 0-10 large numbers						----	.125	-.106
7. Arithmetic Time 1							----	.335**
8. Arithmetic Time 5								----

*Notes. Pearson product-moment correlation coefficient. * p < .05. ** p < .01*

8.3 Conclusion.

Overall, three findings emerged: Children's accuracy as well as linear and logarithmic function fits in number estimation significantly improved over the 25-months' time span from nursery to the end of Year One. Second, it seems that pre-school children really struggled with the number estimation tasks. At Times 1 and 3, most children (78.1% and 79.1% respectively) had a mean error rate of 20%. In contrast, at Time 5, less than half of the children (44.3%) had a mean error rate of 20%. This pattern may be due to children's lack of exposure to number lines before they enter school. Last but not least, the current study showed a weak relationship between early arithmetic and number estimation. Multiple regression analyses showed that counting, and especially transcoding are stronger concurrent predictors of arithmetic than numerical estimation. SEM path models revealed that the 0-20 scale was problematic to model and the 0-10 scale at Time 1 was best described as two distinct factors: small numbers within children's subitizing range (numbers 1-4 on the 0-10 scale) and large numbers on the 0-10 scale. The large-numbers-factor was a significant predictor of arithmetic 25 months later, but transcoding was the strongest predictor. It should be noted that the SEM model as a whole did not provide a good fit to the data.

The results overall support previous findings of number estimation in young children (Muldoon et al., 2013). The results confirmed that children's estimates on the 0-20 scale were better fit by a logarithmic function. Muldoon and colleagues further reported a significant correlation between children's maths achievement and their linear and logarithmic function fits on the scales 0-20 and 0-100 at five years and four months. This study also found that the linear as well as the logarithmic function fits were significantly related to arithmetic at Time 3 (comparable to Time 1 in Muldoon et al., 2013), but not Time 5. However, neither linear nor logarithmic function fit on the 0-10 scale at pre-school age (Time 1) were correlated to arithmetic. It must be noted that pre-school children in this study were not assessed on either scale 0-20 nor 0-100 at Time1 because of the unfamiliarity of the task (there were high error rates and low linear and logarithmic function fits on the 0-10 scale). Pre-school children in the current study had no exposure to number lines. Descriptive statistics showed that though linear and logarithmic function fits at Time 3 may be comparable to Muldoon et al. (2013), error rates were much higher suggesting that even children in the first year of formal schooling may find number estimation too difficult.

Children's counting scores were significantly associated with arithmetic scores at the three time points. The significant relation between counting and arithmetic (Muldoon et al., 2013), was not found once children's performance on transcoding tasks was taken into account. Neither linear nor logarithmic function fits on the scales 0-10 and 0-20 were significant predictors of arithmetic. In contrast, Muldoon et al. (2013) found a significant contribution from linear function fits on the 0-20 scale at five years of age and linear function fits on the 0-100 scale at six years.

In contrast to the findings in the literature that children's estimation on the 0-20 and 0-100 scales are predictive of arithmetic scores, the longitudinal SEM path models showed that, besides transcoding (strongest predictor), four-year-olds estimation on large numbers on the 0-10 scale (numbers '6' and '8') was a significant unique predictor of arithmetic 25-months later. However, there were issues when modelling the number estimation data. First, estimations on the 0-10 formed two factors (small numbers in the subitizing range and large numbers) and second, not all large numbers were significantly loading onto the hypothesised factor ('7' and '9' had to be excluded) which raises the question whether the drawn conclusions on number estimation affecting arithmetic may be invalid, especially considering the poor performance of pre-school children on number estimation. The fact that the model provided a poor fit and the nonsignificant correlations of number estimation to other factors, and between both number line factors in particular, limits the interpretation.

Muldoon's study showed that children's estimation on the 0-100 scale is particularly important when it comes to math achievement. As this study did not assess number estimation on the 0-100 scale, little can be said about this link between number line estimation and arithmetic. However, given the struggle the current sample already had with estimation on the 0-10 and 0-20 scales, it seems unlikely that they would perform with any accuracy on the 0-100 scale. Current findings suggest that understanding of the number line task is limited in pre-school children, and that other factors are driving the development of arithmetic at this age.

Chapter 9. General Discussion

This thesis aimed to explore the relationship between early arithmetic skills (approximate and exact arithmetic skills) and ANS, language and numeracy skills in children transitioning from pre-school to their first years of formal schooling. Four year-old children were assessed on a comprehensive test battery including measures of magnitude comparison, language comprehension and number knowledge (counting, number estimation, number reading, writing and identification) as well as background measures of cognitive skills and inhibition in particular and re-tested four times over a 25-month period.

First, the results showed that children's ability to translate between verbal number codes and Arabic numerals (transcoding) was the only unique significant predictor of children's arithmetic skills two years later. Second, the relationship between symbolic and nonsymbolic magnitude comparison shifts from a two-factor structure in younger children to a unitary, general magnitude comparison factor in Year One. Third, symbolic magnitude comparison at four years was a significant predictor of both symbolic and nonsymbolic approximate arithmetic at five years, six months. Children's performance on transcoding tasks was a significant predictor of symbolic approximate arithmetic and nonverbal intelligence at four years was a significant predictor of nonsymbolic approximate arithmetic. Both symbolic and nonsymbolic approximate arithmetic assessed at six years, four months was significantly predicted by transcoding scores of four year-olds.

9.1 Concurrent and longitudinal relations between arithmetic and transcoding, ANS, counting, number estimation, language, nonverbal intelligence and inhibition skills.

Overall, the following findings emerge:

Children's ability to translate between verbal number codes and Arabic numeral (transcoding; measured as number reading, writing and identification) was a significant concurrent predictor of arithmetic at four years (Time 1), five years; ten months (Time 4) and six years (Time 5). These results confirm and extend Göbel et al.'s (2014) finding that transcoding was a unique longitudinal predictor of six-year-olds' performance on arithmetic. It was noted that the contribution of transcoding at Time 4 was only marginally significant (*p* = .063) which suggests that it may not substantially contribute to explaining the variance in arithmetic at that particular stage. Nonetheless, these findings support the hypothesis that the ability to translate between Arabic numeral and their verbal code crucially impact the development of early arithmetic.

Moreover, a stable pattern emerged when investigating longitudinal prediction of arithmetic. Only transcoding at Time 1 was uniquely predicting arithmetic skills after 9-months, 16-months, 20-months and even 25-months. Few studies have followed children's mathematical development at this early stage. Jordan et al. (2009) investigated the relation between early number competence and mathematics achievement from beginning of kindergarten to the middle of grade 1 showing that early number competence predicted the rate of growth in mathematics achievement. However, their number competency factor comprised of a wide range of numerical skills including counting and number recognition, number comparisons, nonverbal calculation, story problems, and number combinations.

There is ample evidence for the developmental importance of ANS in arithmetic, and the possibility that ANS drives exact arithmetic via approximate arithmetic skills (Halberda et al., 2008; Gilmore et al. 2010; Libertus et al., 2011; Mazzocco et al., 2011). Previous analyses of concurrent associations with arithmetic found that symbolic magnitude comparison in particular was a significant predictor of early arithmetic (Holloway and Ansari, 2009; de Smedt et al., 2013; Siegler, 2016). In the present study, as children moved to formal schooling, symbolic and nonsymbolic comparison tasks loaded significantly onto a unitary factor rather than two distinct factors. The general magnitude comparison factor was a significant concurrent predictor of arithmetic at five years, ten months and six years. It seems that formal magnitude representations may be crucial at particular stages in children's arithmetic development. However, when looking at the longitudinal prediction of ANS, this study did not find a significant contribution to children's performance on arithmetic tasks assessed later on, , though strongly correlating with later arithmetic skills, confirming Göbel et al. (2014). It was noted that, contrary to Göbel et al. (2014), the symbolic and nonsymbolic magnitude comparison tasks at four years comprised of two separate latent factors. This questions previous findings concerning the central role of the ANS in early arithmetic development (e.g., Piazza, 2010). Further studies are necessary to clarify the longitudinal role of the ANS in the development of early arithmetic and whether the present results hold stable for longer follow-up.

Gilmore and colleagues argue that relationships between nonsymbolic magnitude comparison and arithmetic may be driven by inhibition skills. Contrary to the findings of Gilmore et al. (2013) showing that children's inhibitory control predicted arithmetic after controlling for performance on dot comparison, this study showed the reverse pattern that only dot comparison (ANS) predicted children's arithmetic even after taking inhibition into account. It is worth mentioning that the sample in Gilmore et al. (2013) were older and, despite the broad age range (seven to ten years), the analyses did not control for age. Their measure of inhibition was the NEPS-II inhibition subtask (Korkman, Kirk, and Kemp, 1998), a GoNoGo test, similar to our Head-Toes-Shoulders-Knees task. Children's performance on inhibition tasks did not explain variance of arithmetic scores once performance on nonsymbolic magnitude comparison has been accounted for.

In regards to counting, we found that counting was unique predictor of pre-school children's arithmetic performance. Indeed, researchers have reported that counting is important for calculating (Ansari et al., 2003; Cowan et al, 2005). Furthermore, Donlan et al. (2007) found a strong association between counting and calculation suggesting that the performance on both tasks draw from a common representational system.

However, the current results reveal that though concurrently predicting arithmetic, children's performance on counting tasks was not a significant longitudinal predictor of arithmetic as suggested in some models of arithmetic development (Aunola et al., 2004; Zhang et al., 2014 and LeFevre et al., 2010). Aunola et al. (2004) measured counting in a more complex way than in the current study, including rote counting, counting forwards and backwards from given number and counting in steps. Of particular interest is the fact that Aunola et al. (2004) included a number identification task (similar to measure used in the current study) as part of the outcome measure of maths. It could be that the administration of a broader, more complex counting task as well as the difference in designing the SEM path model (number identification was part of the outcome math measure in Aunola et al. (2004) compared to being a predictor in the current path models) may be reason for the contrasting results.

In regards to the contribution of children's number estimation, the longitudinal SEM CFAs showed that four-year-olds number line estimation was best described by two factors: performance on 0-10 scale small numbers (subitizing range) and 0-10 scale large numbers. The results showed that transcoding (strongest predictor) and children's performance on large number not within the subitizing range (numbers '6' and '8') were uniquely predicting arithmetic 25-months later. However, there were issues when modelling the number estimation data which raises the question whether the drawn conclusions on number estimation affecting arithmetic may be void considering the poor performance of pre-school children. The fact that the model provided a poor fit limits the interpretation.

Evidence for the importance of language in the development of arithmetic comes from studies assessing the mathematical skills of children with SLI which performed lower in rote counting than typically developing children of the same age (Donlan et al., 2007). SLI may affect a wide range of numeracy skills differently (Donlan, Bishop and Hitch, 1998; Donlan, and Gourlay, 1999; Fazio, 1994, 1996), but the relationship between these skills is complex, and runs counter to other findings which indicate independence between verbal and nonverbal calculation skills (Nunes and Bryant, 1996; Jordan et al., 1994).

This thesis did not find a significant relationship between language comprehension in general and early arithmetic, neither cross-sectional nor longitudinal. However, the study found significant, concurrent relationship between math-related language comprehension and arithmetic at four years, eleven months as well as longitudinally. Math-related language comprehension assessed at four years was a significant predictor of arithmetic skills assessed nine months later, supporting the notion that language impacts early arithmetic (Donlan et al., 1998; Donlan, and Gourlay, 1999; Fazio, 1994; 1996; Donlan et al., 2007; Cowan et al., 2005; Kleemans et al., 2011; 2012). However, most studies assessed language skills by more general language tasks neglecting language specific to mathematics. It appears that not general language comprehension but math-related language comprehension, such as the understanding of *more*, may be a key foundation in acquiring arithmetic.

In regards to nonverbal ability, the current study found that children's nonverbal intelligence was the strongest significant predictor of arithmetic tasks at age of 4 years, 3 months (Time 1) supporting the claim that general intelligence affects children’s early arithmetic skills (Cowan et al., 2005; Noël, 2009). However, this was not confirmed for the longitudinal prediction of arithmetic. Also, findings from studying number wizards (high achievers on the number reading task) showed that the importance of nonverbal ability in pre-schoolers drops once children’s number recognition (number reading task) was taken into account. This could be the subject of further study. The findings, however, support the idea that knowing your numbers is crucial for the development of early arithmetic and that children who have already mastered the Arabic numerals from one to ten may not rely on cognitive processes such as nonverbal intelligence but rather engage specialised math-related skills such as math-related language and transcoding (numerical knowledge).

Last but not least, it is important to take account of the finding that neither of the arithmetic autoregressor (Times 1 or 2 arithmetic performance) significantly predicted arithmetic at later stages. This could be due to the fact that the testing procedure for early arithmetic was changed over the course of the study to adjust for children’s growing learning experience in arithmetic. In particular, there were ceiling effects with arithmetic scores assessed at Time 1 (children could complete the test in their own time). At Time 2, time to solve arithmetic was limited to three minutes, but the time limit may have been too long to achieve high sensitivity at this age. It seems that the TOBANS, used in later testing, was a sensitive measure of arithmetic and it would be interesting to see if children as young as four years can successfully perform on the TOBANS.

9.2 Development of magnitude comparison

Overall, the comprehensive analyses of children’s performance on symbolic and nonsymbolic magnitude comparison tasks revealed three findings: First, children’s performance on magnitude comparison tasks generally showed significant distance effects for both symbolic and nonsymbolic comparisons with better performance on the far than close trials confirming previous findings (Barth et al., 2003; Piazza et al., 2010; Xu and Spelke, 2000; Halberda et al., 2008; Gilmore et al. 2010; Libertus et al., 2011; Mazzocco et al., 2011; Halberda and Feigenson, 2008). However, the symbolic distance effect was not evident at Time 1. Some children had difficulties reading the Arabic numerals at Time 1. A third of the children made at least two mistakes in reading the single digit Arabic numerals 0-9. After taking mastery of the single digit Arabic numerals into account, a marginal distance effect was found even in young children. As expected, no such limitations applied to performance on nonsymbolic comparison.

Second, the results revealed ratio effects (2:3 > 3:4 > 5:6) across all time points. Children performed more accurately on ratios with a large difference (i.e. 2:3) than ratios with a small difference (5:6). The feature size (fixed size versus surface-area matched) impacted children's performance on comparison tasks. Fixed size arrays were easier to discriminate for both, distance and ratio trials, than surface-area matched arrays suggesting that it is more difficult for children to ignore the prominent feature size in the surface-area matched condition where the array with fewer stimuli has bigger squares compared to many tiny squares. The results showed an increase in performance over time with six year-olds performing almost adult-like confirming Halberda and Feigenson (2008).

Third, the relation between symbolic and nonsymbolic magnitude comparison tasks is dynamic and changes over time. The change from a two-factor model (symbolic and nonsymbolic) to a single-factor model occurs with children's entry to the formal school system. Symbolic and nonsymbolic comparison tasks loaded on separate factors at pre-school age confirming Libertus et al. (2011), Piazza (2010) and Piazza and Dehaene (2004). The distinction between the factors is declining over time and it seems that the children's representation and processing of magnitude comparison tasks at the age of 5 years, 6 months (autumn term of Year One) is changing towards the general comparison ability construct. At Times 4 and 5, the single-factor model was the best fitting model meaning that magnitude comparison tasks load onto one general comparison factor and not two distinct factors confirming the developmental trend towards a general magnitude comparison factors (see also Kolkman et al., 2013; Göbel et al., 2014). Questions remain on why this change in the representation and processing of the magnitude comparison occurs.

9.3 Performance and structure of early approximate arithmetic skills

With regards to children's performance on approximate arithmetic tasks, the results replicated the finding that young children can perform nonsymbolic as well as double-digit symbolic approximate arithmetic with accuracy above chance, and that nonsymbolic approximate arithmetic was easier for young children than symbolic approximate arithmetic (Gilmore et al., 2007). The young children tested in this thesis were not as accurately as reported in Gilmore et al. (2007) and only children at

Times 4 and 5 (six years and over) showed comparable performance. The results further revealed ratio effects (4:7 > 4:6 > 4:5) for symbolic and nonsymbolic approximate arithmetic across all three time points. Children performed more accurately on ratios with a large difference (4:7) than ratios with a small difference (4:5) similar to Gilmore et al. (2007).

Symbolic approximate arithmetic showed higher correlations with exact arithmetic compared to nonsymbolic approximate arithmetic suggesting that symbolic approximate arithmetic may be more closely related to conventional exact arithmetic tasks. It must be critically mentioned that nonsymbolic approximate arithmetic was only weakly related to exact arithmetic, and symbolic approximate arithmetic correlated only moderately. To what extent these correlations represent common processes in the performance of approximate and exact arithmetic is still unclear. Might the moderate correlation between symbolic approximate arithmetic and exact arithmetic tasks be explained by common demand on symbol identification? Note in general that these are zero-order correlations which allow only limited interpretation.

Some tentative conclusions may be drawn when investigation to what extent language, ANS and transcoding may impact approximate arithmetic. It seems that symbolic magnitude comparison at Time 1 is crucial for children's understanding of early approximate arithmetic, whether symbolic or nonsymbolic. Why would that be? Easier to interpret is the involvement transcoding, which operates as a consistent longitudinal predictor of both symbolic and nonsymbolic approximate arithmetic in six year olds. This result strengthens the finding of transcoding as the only unique longitudinal predictor of exact arithmetic. To what extent does knowledge of the symbol system underpin arithmetic, and to what extent can the relationships observed here explained by shared cognitive processes, such as magnitude comparison.

Symbolic and nonsymbolic approximate arithmetic tasks are best explained by independent factors. This relationship does not change over time contrary to the development of magnitude comparison. Does this switch from the two-factor towards the one-factor model occur later, perhaps in the second full year of formal schooling? Or maybe symbolic and nonsymbolic approximate arithmetic may never form one factor due to the nature of the tasks being so distinct. Further research is needed to clarify the structure of symbolic and nonsymbolic approximate arithmetic as well as their relation to exact arithmetic.

9.4 Children's performance on number estimation line tasks

Children's accuracy as well as linear and logarithmic function fits on a number estimation task significantly improved over the 25-months time span from nursery to the end of Year One supporting Muldoon et al. (2013). It seems that pre-school children struggled with the number estimation tasks which may be explained by children's lack of exposure to number lines before they enter school. The reported significant relation between counting and arithmetic (Muldoon et al., 2013), was not found once children's performance on transcoding tasks was taken into account. Neither linear nor logarithmic function fits on the scales 0-10 and 0-20 were significant predictors of arithmetic.

Muldoon's study showed that children's estimation on the 0-100 scale is particularly important when it comes to math achievement. This study did not assess number estimation on the 0-100 scale because of the unfamiliarity of the task which was supported by children's high error rates and low linear and logarithmic function fits on the 0-10 scale. It would be of interest to establish pre-school children's performance on number estimation on the 0-100 scale. Nonetheless, it can be speculated that young children most likely fail to produce sensible data considering the struggle the current sample already had with estimation on the 0-10 and 0-20 scales. Results of young children and pre-school children in particular will be invaluable when it comes to the understanding of the development and linearity of the number estimation task.

9.5 Implications for models of mathematical development

Based on Dehaene's Triple Code (1992) it can be hypothesised that arithmetic at a young age may draw on the magnitude code (ANS). This analogue magnitude code is involved in the direct route for solving problems and is proposed to be used in subtractions and more complex addition problems. This route seems to be the most applicable to the early stage of arithmetic. The pre-school children assessed in this study were not formally trained on arithmetic, it can be assumed that even simple additions as used at Time 1 are as challenging and difficult as complex additions are for older children. Interestingly, the current results found that ANS, as proposed by the Triple Code (1992), is not the primary driver of early arithmetic development. Rather children's ability to translate between the verbal and written form of Arabic numerals was the primary driver of the development.

These findings bear some very interesting implications to the developmental models of arithmetic of children's problem solution discussed in the opening chapter, especially the Four-Step Developmental Model of Numerical Cognition (von Aster and Shaley, 2007), the Pathway Model (LeFevre et al., 2010) and the Integrated Theory of Numerical Development (Siegler and Braithwaite, 2017).

The Four-Step Developmental Model of Numerical Cognition (von Aster & Shalev, 2007) proposes that children move through four stages as they progress in arithmetic through exposure and an increase in working memory capacity (Figure 1.6). The inherited core-system of magnitude representation (first stage) entails subitizing and approximation abilities. The finding that pre-school children score highly on an individual administered non-time-constrained magnitude comparison task, confirms the model's first assumption, that children learn the first core system of magnitude in infancy.

Of particular interest for our findings are steps 2 and 3 (pre-school and school age, respectively) as this thesis investigated this sensitive period of transitioning from pre-school to school stage. According to the model, pre-school children move on to the linguistic stage of numeracy (step 2) where children acquire the verbal number codes (counting). In step 3, children learn the Arabic number system and the symbolic representations of magnitudes in school. Typical mathematical skills developing at this stage are written calculations and odd-even decisions.

The results showed that pre-school children acquired the counting words as predicted by the model, however pre-schoolers have not fully mastered the number word sequence making mistakes in the teen-numbers. Only few could count up to 30 or above. The results further showed that the sample performed above chance level on tasks assessing Arabic number system (transcoding tasks) and written additions using counting strategies supporting step 2 of the model. Further evidence for step 2 comes from the concurrent associations between counting and arithmetic in pre-schoolers' confirming that children learn to count before entering formal schooling. However, the current sample could also solve subtractions at the beginning of Year One which suggests that their arithmetic skills may be more advanced than proposed by Von Aster and Shalev (2007). Subtractions at this age can be considered as a more complex arithmetic skills because children have not been formally taught neither additions nor subtractions.

In the four-step-developmental model, acquisition of the Arabic number system such as the digits. Contrary to the model, our analyses found that pre-school children have already a basic understanding of the Arabic number system (transcoding tasks, especially number identification). Four-year-old children performed above chance level on the number identification task, including double digits. The results showed that this basic Arabic number system knowledge, preceding formal arithmetic training, plays a crucial role in later exact arithmetic.

The final stage of the model, understanding and manipulation of the mental number line, develops during school years as children acquire the concept of ordinality, a second core principal of number. The model posits approximate arithmetic skills and mental number line estimations at this later stage. The current findings confirmed that pre-school children as well as Year One children struggled on number estimation tasks, however, we found that approximate arithmetic has proven useful when testing pre-school children who have an immature knowledge of exact arithmetic and had no formal training of exact arithmetic, yet are able to perform above chance level on approximate arithmetic, contradicting the model (approximate arithmetic is supposed to emerge later in school). It may be that children’s early approximate arithmetic skills foster later exact arithmetic abilities.

To sum up, in accordance with the model, the current findings support the idea that children learn magnitude comparison early in life, followed by counting in pre-school moving on to additions, more complex calculations in school and the mental number line. In contrast to the model, the results suggest that children’s understanding of the Arabic number system develops in pre-school, hence earlier than proposed in the model. Also, it seems that approximate arithmetic may be acquired in the first school years and not later on.

A second important developmental model is the Pathways to Mathematics model (LeFevre et al., 2010; Figure 1.7) focusing on the relationships between children's cognitive precursors of early numeracy skills and mathematical outcomes. This model posits three separate pathways: quantitative, linguistic, and spatial attentional. Each of these pathways contributes individually to the acquisition of early numeracy abilities. Furthermore, the models proposes that the linguistic, quantitative and spatial attention pathways vary in their contribution to mathematical performance depending on the demands of the arithmetic problem. Only the linguistic and quantitative pathway are important to this thesis. According to the model, linguistic skills are directly linked to children's symbolic number system knowledge which is further linked to geometry, numeration skills (number line and calculation) and magnitude comparison. Quantitative skills are related to processing numerical magnitudes and magnitude comparison.

LeFevre and colleagues only assessed general receptive language comprehension (vocabulary test) and phonological awareness to capture linguistic precursors. In this thesis both, general language comprehension as well as math-related language were assessed. The results showed that math-related language, not language comprehension in general, was a significant predictor of arithmetic (concurrent analyses at Times 2 and 3; longitudinal prediction of arithmetic 9-months later). Contrary to the model, linguistic skills affected arithmetic directly and not via children's symbolic number system (number naming task similar to the one included in transcoding factor) as proposed in the model. This further shows that it is important to analyse the distinct contributions of general language comprehension and math-related language. Nonetheless, the link between symbolic number system (similar to the transcoding factor) and arithmetic was confirmed in this study because children's transcoding skills were the only stable and unique longitudinal predictor of arithmetic performance.

The Pathways model assessed quantitative skills through children's subitizing skills. This study showed that symbolic magnitude comparison was a significant concurrent precursor of arithmetic at Time 3 supporting the quantitative pathway. Symbolic and nonsymbolic magnitude comparison gradually integrated to a general magnitude comparison factor (five years, six months of age). The results further showed significant associations between the integrated, general magnitude factor and arithmetic in Year One (Times 4 and 5). However, magnitude comparison performance was not a significant longitudinal predictor once children's understanding of the Arabic number system (transcoding) was taken into account, suggesting a particular driving force in number symbol knowledge.

To sum up, contrary to the model the results suggest a direct linguistic pathway between math-related language comprehension and arithmetic. It is important that future models should distinguish between children's general language comprehension and language related to mathematics. Furthermore, the analyses confirmed the quantitative pathway. No statements can be made of the spatial pathways because the current thesis did not measure any spatial skills. Because of the sample size, we only investigated direct prediction of children's arithmetic skills, but not complex mediation models.

The Integrated Theory of Numerical Development (Siegler and Braithwaite, 2017) proposes that numerical magnitudes are represented on a mental number line, providing the basis for mathematical activity. Smaller numbers are presented on the left and larger number on the right of the number line. The representation of whole numbers shifts from a logarithmic distribution towards a linear distribution during the primary school years. This shift occurs first for small whole numbers than larger whole numbers based on children's experience with the number range. Siegler and Braithwaite (2017) posit that numerical magnitude knowledge is related to and predictive of arithmetic development. The model is very sparse. It does not identify factors or systems which might drive numerical development, and does not differentiate, for example, spoken versus written number forms.

This model states that the development of small whole numbers on a 0-10 number line occurs in pre-school (three to five years) and that school children (five to seven years) can solve number line estimations on a 0-100 scale. Our results contradict these assumptions showing that pre-school children struggled on number estimation tasks on the 0-10 scale. Children's error rates were very high in pre-schoolers' and were only acceptable at the end of Year One. Regarding the proposition that mental number representation is predictive of arithmetic, the results found that only number estimations for the numbers '6' and '8' were weakly predictive of arithmetic skills 25-months later, while children's transcoding skills proved to be a more powerful and consistent predictor of arithmetic.

Some studies propose that, rather than being a precursor of mathematical achievement, number line acuity and math performance both influence each other during development from pre-school through early school years (Friso-van den Bos, Kroesbergen, Van Luit, Xenidou-Dervou, Jonkman, Van der Schoot and Van

Lieshout, 2014; LeFevre, Lira, Sowinski, Cankaya, Kamawar and Skwarchuk, 2013). The relation between number estimation and arithmetic, particularly in pre-school and primary school, needs to be further studied, in the context of other factors, especially transcoding.

Furthermore, children's number estimation on the 0-10 scale showed a more linear distribution at all testing points contradicting Siegler's proposition that mental representation shifts from a logarithmic to a more linear distribution. Children's number estimates for the 0-20 scale showed similar linear and logarithmic distributions. Either mental number representation does not undergo this shift from logarithmic to linear, and these representations may be even more complex than proposed by Siegler, or this shift happened before the testing period started. However, the latter seems improbable because neither the model nor our accuracy data suggest that children younger than four years have an established mental number line for small whole numbers.

To sum up, the findings suggest that the model overestimates the age at which a mental number representation is established. Our results are more in line with the four-step-developmental-model supporting the idea that children acquire an understanding of the mental number line later on in school.

9.6 Method Constraints.

It must be critically mentioned that not all measures were assessed at all testing points, thus constraining conclusions drawn about the concurrent prediction of arithmetic and the change of the relationships with arithmetic over time. Also, the testing procedure of tasks was changed to adjust for children's growing learning experience in the tasks measured (see assessment of magnitude comparison and arithmetic for more details). Further studies are needed to investigate the concurrent prediction using the same tasks at all testing points to enable comprehensive conclusions about prediction of early arithmetic and how the concurrent relationships may change over time.

Also, the Time 1 counting measure (highest number produced in correct order) produced high variability in scores. A more complex measure such as the composite of various counting tasks assessing a wider range of counting skills used by Aunola et al. (2004) may prove to be useful in examining the relationship between ANS, counting, transcoding and arithmetic.

A further constraint of the study was the relatively small sample size which made it impractical to investigate more complex SEM path models including additional covariates of math achievement such as children's early memory or spatial skills. Further large scale studies are needed to clarify longitudinal prediction using more sensitive and comprehensive measures to enable more detailed conclusions about the prediction of early arithmetic, and to ascertain whether the findings hold stable for longer-term follow-up.

9.7 Future Directions.

This thesis has focused only on ANS, language and numerical knowledge as predictors of arithmetic. There is widespread evidence that other factors might be important too. Other abilities, than those investigated in this thesis, that have been reported to be important in children's arithmetic development are aspects of working memory, such as central executive functioning (Bull et al., 2011), phonological loop (Swanson and Sachse-Lee, 2001; Wilson and Swanson, 2001) and the visuo-spatial sketchpad (Bull et al., 2008; Rasmussen and Bisanz, 2005; Holmes, Adams, and Hamilton, 2008; LeFevre et al., 2011) and phonological awareness (Baldo and Dronkers, 2007; Dehaene et al., 2003).

Another point to consider is children's early number estimation ability. There is evidence (Muldoon et al., 2013) that young children (five years-olds) accurately place numbers on a horizontal 0-10, 0-20 or 0-100 number line. However, this study found that young children, and pre-school children without any formal teaching on arithmetic or exposure to number lines, struggled considerably with the 0-10 and 0-20 scaled trials questioning the idea to assess pre-school number estimation. Simms et al. (2013) showed that four- to seven-years old children displayed numbers as accurately in the vertical as the horizontal orientation. It would be of interest to explore pre-school children's number estimation on vertical number lines and compare the performance with horizontal number line. It is possible that vertical number lines may be easier because children in pre-school have much more exposure to stacking bricks and blocks to build towers, counting the objects as they go. These observations highlight the role of input and experience in forming early representations of number, and the importance of including these measures in future studies.

Of particular interest for future research is the finding that literacy and transcoding factors at six years were highly correlated (Table 4.6). An important correlate of maths disability is reading disability. It is estimated that 40% of dyslexics also have maths disability (Lewis, Hitch, and Walker, 1994). It is still debated whether there are cognitive processes that are shared between the two learning disorders or whether dyslexia and dyscalculia are largely independent on a cognitive level. The finding that children’s ability to translate between the spoken and written form of numbers is a powerful longitudinal predictor of arithmetic skills in primary school (current study; Göbel et al., 2014) suggests that children’s knowledge of Arabic numerals may represent a critical foundational skill underlying early arithmetic, analogous to the role of letter knowledge in reading, and may be crucial for further arithmetic development. Further research is needed to establish whether reading and arithmetic share common cognitive processes supporting the association between written and their spoken referents.

9.8 Conclusion

The scope of this thesis was to investigating predictors of arithmetic skills at five time points from pre-school through to the end of the first year of formal schooling.

Regarding the concurrent associations, transcoding, children’s ability to translate between spoken and symbolic form of numbers, was shown to play an important role in the development of early arithmetic skills in one form or another at every testing point. Transcoding may not have been the strongest or only predictor at times, and other factors may also impact the development of early arithmetic at different time points. At an early, pre-school age, it appears that nonverbal intelligence, counting and math-related language, particularly children’s understanding of *more*, in addition to transcoding affect the performance on arithmetic tasks. These relations however, weaken in favour of the relationship with magnitude comparison in early school years (Year One). After children entered the formal schooling system, both transcoding and a general magnitude comparison factors were crucial for arithmetic development.

Cross-sectional designs only take snapshots of development at different stages. There are limitations to the conclusions that can be drawn, since the measurements taken differ at different time points. Nonetheless, the results are informative concerning the process of development, and provide a useful background to the research exploring longitudinal prediction of arithmetic. Although cross-sectional relationships may draw attention towards special relations between arithmetic and its precursors, it is only one way of investigating the development of arithmetic skills.

As for longitudinal prediction, a stable pattern emerged that only four-year-olds' transcoding ability was uniquely predicting arithmetic skills after 9-months, 16-months, 20-months and even 25-months. This finding extends the findings of previous studies (Göbel et al., 2014; Jordan et al., 2009) and challenges the proposal that the approximate number system at pre-school age drives the development of arithmetic (Mazzocco et al., 20011). Children's ability to translate between spoken form and Arabic numeral relies on Arabic-digit knowledge and place-value understanding. Children's understanding of place-value may be a key foundation for the development of later arithmetic skills (Möller et al., 2011), as may Arabic-digit knowledge at school entry (Kolkman et al., 2013; Krajewski and Schneider, 2009; Mundy and Gilmore, 2009). The latter relationship appears to be directly analogous to the critical longitudinal role of early letter knowledge on the development of reading skills (Caravolas et al., 2012; Hulme et al., 2012). In short, the results suggest that learning arithmetic may share some developmental pathways with learning to read. Learning the symbol set (Arabic numerals or letters) and their verbal labels is a critical foundational skill of later arithmetic skills.

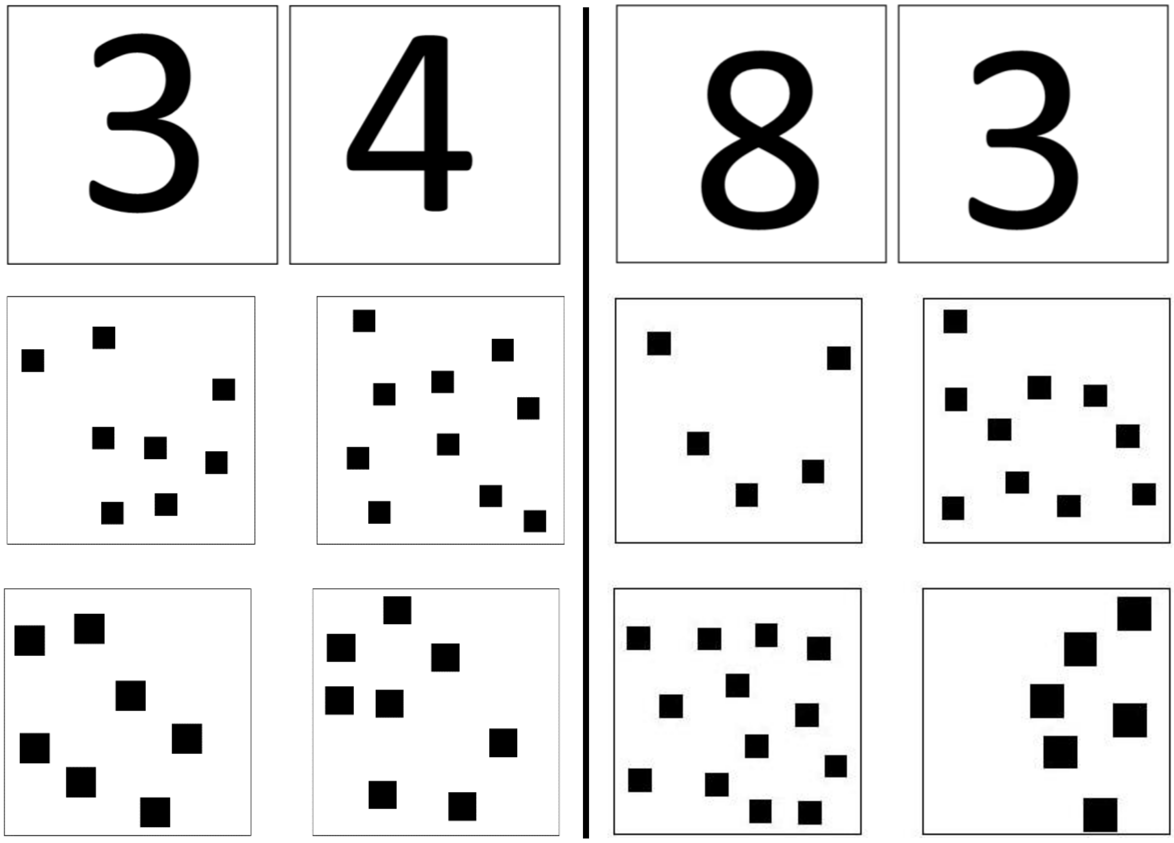
The thesis further clarified the little investigated developmental structure of symbolic and nonsymbolic magnitude comparison. At pre-school age, ANS tasks describe two distinguishable skills compared to the later integration of ANS skills into one general magnitude comparison structure. In view of these recent results, previous findings on the relationship between symbolic and nonsymbolic comparisons and their impact on arithmetic skills at school age should carefully be re-examined.

Furthermore, the symbolic approximate arithmetic task introduced by Gilmore et al. (2007) has proven useful when testing pre-school children who have an immature knowledge of exact arithmetic and had no formal training of exact arithmetic. However, detailed analyses do not support the statement made by Gilmore et al. (2007) that children 'used nonsymbolic number representations to solve symbolic problems' (p.590). Longitudinal analyses showed a common dependence on symbolic comparison and a specific contribution of nonverbal ability to nonsymbolic approximate arithmetic, which is then superceded by a common influence of early symbol transcoding. Future research may explore whether there will be a later shift, similar to magnitude comparison, from a two-factor model with symbolic and nonsymbolic approximate arithmetic as distinct constructs towards a unitary model with one general approximate arithmetic construct.

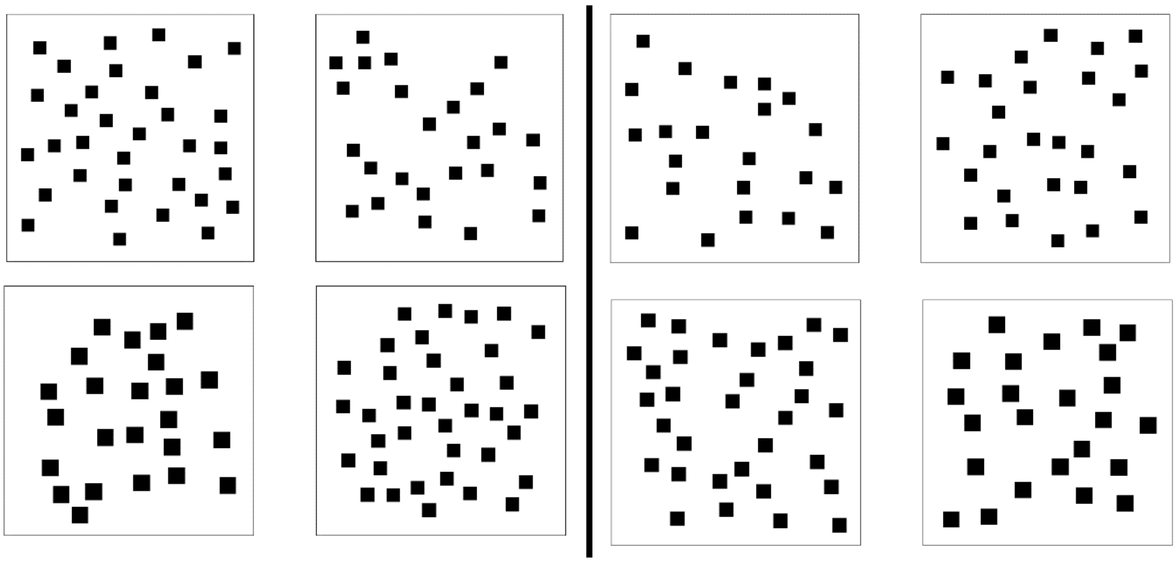
It is hoped that the detailed findings summarized above, emphasizing the importance of number symbol knowledge, will prove useful both in enhancing the educational experience of young children learning about numbers, and in identifying and supporting those who struggle.

Appendices

Appendix 1: Symbolic and Nonsymbolic Magnitude Comparison Tasks



Example symbolic (above) and nonsymbolic (fixed size in the middle and surface-area matched below) distance effect comparison (close items are on the left and far items on the right).



Example ratio effects for fixed size (above; 2:3 on the left and 3:4 on the right) and surface-area matched (below; 3:4 on the left and 5:6 on the right).

Appendix 2: Test of Relation Concepts (TRC) at Time 1

Part 1 - Easy

	Practice items	Correct Response	Response Given
A	The boy has more carrots than the girl	1	
B	The brown sheep has more wool than the grey sheep	3	
C	The purple witch is more beautiful than the yellow witch	2	

1	The girl has more chocolate than the boy	1	
2	The boy has more pasta than the girl	3	
3	The boy has more bananas than the girl	3	
4	The girl has more flowers than the boy	4	
5	The boy has more ice-cream than the girl	2	
6	The girl has more milk than the boy	4	
7	The girl is more colourful than the boy	1	
8	The yellow bed is more comfortable than the blue bed	4	
9	The girl has more balloons than the boy	1	
10	The boy has more biscuits than the boy	2	
11	The short dress is more colourful than the long dress	3	
12	The green chair is more comfortable than the red chair	2	

/12

Part 2 - Hard

	Practice items	Correct Response	Response Given
A	The boy has more carrots than the girl	1	
B	The brown sheep has more wool than the grey sheep	3	
C	The purple witch is more beautiful than the yellow witch	2	

1	The pink man is more handsome than the blue man	2	
2	The red house has more smoke than the blue house	3	
3	The girl has more chips than the boy	3	
4	The girl has more butter than the boy	3	
5	The yellow planet has more aliens than the yellow planet	2	
6	The girl has more cheese than the boy	1	
7	The boy has more apples than the girl	3	
8	The green prince is more handsome than the blue prince	4	
9	The girl has more eggs than the boy	1	
10	The red baby is more beautiful than the blue baby	1	
11	The boy has more bread than the girl	4	
12	The blue princess is more beautiful than the green princess	4	

/12

Appendix 3: Test of Relation Concepts (TRC) more and less (Times 2, 3, 4 and 5)

Practice items		Correct Response	Response Given
a	the boy has more carrots than the girl	4	
b	the brown sheep has less wool than the grey sheep	3	
c	the purple witch is more beautiful than the yellow witch	3	

1	the girl has more chocolate than the boy	1	
2	the boy has less pasta than the girl	3	
3	the boy has more bananas than the girl	2	
4	the girl has fewer flowers than the boy	2	
5	the red baby is less beautiful than the blue baby	4	
6	the boy has more ice-cream than the girl	2	
7	the girl has more milk than the boy	3	
8	the girl is less colourful than the boy	1	
9	the red house has less smoke than the blue house	1	
10	the yellow bed is more comfortable than the blue bed	4	
11	the girl has more balloons than the boy	3	
12	the boy has more biscuits than the girl	1	
13	the long dress is more colourful than the short dress	4	
14	the green chair is less comfortable than the red chair	2	
15	the yellow planet has fewer aliens than the red planet	1	
16	the pink man is more handsome than the blue man	2	
17	the girl has more chips than the boy	4	
18	the girl has more butter than the boy	3	
19	the girl has less cheese than the boy	4	
20	the boy has fewer apples than the girl	2	
21	the green prince is less handsome than the blue prince	4	
22	the blue princess is more beautiful than the green princess	1	
23	the girl has fewer eggs than the boy	3	
24	the boy has less bread than the girl	3	

Appendix 4: Number Identification Task Time 1

11		1	
3	13	8	
15	7	50	5

9	6	8	3
82	208	20	28
206	260	26	2060
706	17	7	70
19	119	91	9
12	22	1	2
41	42	14	4
'[502	25	52	5
1	101	111	11
96	69	6	49
37	13	713	73
7800	807	870	78
43	4	34	304
17	174	74	7
1068	618	18	168
13	3	30	33
58	850	5	85

Appendix 5: Number Identification Task Time 2

11	1		
3	13	8	
15	7	50	5

9	6	8	3
82	208	20	28
206	260	26	2060
706	17	7	70
19	119	91	9
3056	356	35	536
12	22	1	2
41	42	14	4
502	25	52	5
1	101	111	11
96	69	6	49
37	13	713	73
7800	807	870	78
43	4	34	304
17	174	74	7
1068	618	18	168
13	3	30	33
67	670	6710	461
461	14	614	6104
58	850	5	85

Appendix 6: Number Identification Task Time 3

<table>
<tr><td colspan="6">11</td><td colspan="6">1</td></tr>
<tr><td colspan="4">3</td><td colspan="4">13</td><td colspan="4">8</td></tr>
<tr><td colspan="3">15</td><td colspan="3">7</td><td colspan="3">50</td><td colspan="3">5</td></tr>
</table>

9	6	8	3
82	208	20	28
206	260	26	2060
706	17	7	70
19	119	91	9
3056	356	35	536
15	59	50	505
41	42	14	4
502	25	52	5
1	101	111	11
96	69	6	49
37	13	713	73
7800	807	870	78
43	4	34	304
17	174	74	7
1068	618	18	168
13	3	30	33
67	670	6710	461
461	14	614	6104
58	850	5	85

Appendix 7: Number Identification Task Time 4

<table>
<tr><td colspan="2">11</td><td colspan="2">1</td></tr>
<tr><td>3</td><td colspan="2">13</td><td>8</td></tr>
<tr><td>15</td><td>7</td><td>50</td><td>5</td></tr>
</table>

9	6	8	3
82	208	20	28
206	260	26	2060
706	17	7	70
19	119	91	9
3056	356	35	536
15	59	50	505
41	42	14	4
502	25	52	5
1	101	111	11
96	69	6	49
37	13	713	73
3013	10313	313	3030
7800	807	870	78
43	4	34	304
17	174	74	7

1068	618	18	168
13	3	30	33
67	670	6710	461
461	14	614	6104
58	850	5	85
8090	890	89	819
1901	11009	1109	109
617	167	670	6107
11120	1220	100120	120

Appendix 8: Number Identification Task Time 5

11	1		
3	13	8	
15	7	50	5

9	6	8	3
82	208	20	28
206	260	26	2060
706	17	7	70
114	414	440	40014
19	119	91	9
3056	356	35	536
15	59	50	505
41	42	14	4
4017	714	70040	7014
502	25	52	5
1	101	111	11
96	69	6	49
37	13	713	73
582	5028	538	528
3013	10313	313	3030
7800	807	870	78
4708	4807	40087	478
43	4	34	304
17	174	74	7
1068	618	18	168
70031	31	731	713

13	3	30	33
67	670	6710	461
461	14	614	6104
852	925	952	90052
58	850	5	85
8090	890	89	819
1901	11009	1109	109
617	167	670	6107
10212	1220	122	2120
493	943	4930	439

Appendix 9: Arithmetic at Time 1

Practice items	Correct Response	Response Given	Method Used*	
1 + 1	2			
1 + 2	3			

1 + 3	4			
2 + 1	3			
2 + 2	4			
1 + 4	5			
3 + 1	4			
1 + 5	6			
2 + 3	5			
1 + 6	7			
3 + 3	6			
4 + 4	8			

*Method Used: retrieval, counting objects, fingers, guessing etc.

Appendix 10: Arithmetic at Time 2

Form A

Practice items	Correct Response	Response Given	Method Used*	
1 + 1	2			
1 + 2	3			

1 + 3	4			
2 + 1	3			
1 + 5	6			
2 + 3	5			
4 + 5	9			
7 + 2	9			
3 + 5	8			
4 + 2	6			
5 + 2	7			
2 + 6	8			

Form B

Practice items	Correct Response	Response Given	Method Used*	
1 + 1	2			
1 + 2	3			

1 + 4	5			
3 + 1	4			
2 + 5	7			
4 + 2	6			
1 + 6	7			
3 + 6	9			
2 + 7	9			
6 + 2	8			
4 + 3	7			
3 + 5	8			

*Method Used: retrieval, counting objects, fingers, guessing etc.

Appendix 11: TOBANS Simple Addition (Times 3, 4 and 5)

Addition test

2 + 1 =	5 + 4 =
1 + 4 =	2 + 5 =
1 + 2 =	5 + 2 =
4 + 1 =	7 + 2 =
3 + 5 =	4 + 5 =
2 + 3 =	1 + 7 =
1 + 3 =	3 + 3 =
4 + 2 =	4 + 4 =
3 + 1 =	2 + 6 =
2 + 7 =	8 + 1 =
4 + 3 =	5 + 5 =
1 + 5 =	6 + 3 =
6 + 2 =	2 + 2 =
2 + 8 =	5 + 3 =
3 + 6 =	8 + 2 =

Appendix 12: TOBANS Addition with carry (Times 3, 4 and 5)

Addition with carry test

9 + 2 =	3 + 8 =
8 + 4 =	2 + 9 =
5 + 7 =	9 + 8 =
7 + 4 =	7 + 6 =
6 + 5 =	8 + 5 =
9 + 7 =	8 + 7 =
5 + 9 =	6 + 9 =
6 + 8 =	3 + 9 =
6 + 6 =	6 + 7 =
5 + 8 =	9 + 9 =
7 + 8 =	7 + 9 =
6 + 5 =	9 + 3 =
7 + 7 =	6 + 7 =
8 + 9 =	8 + 4 =
4 + 6 =	9 + 7 =

Appendix 13: TOBANS Simple Subtraction (Times 4 and 5)

Subtraction test

2 – 1 =	8 – 5 =
4 – 1 =	9 – 5 =
4 – 2 =	5 – 4 =
7 – 3 =	6 – 2 =
6 – 1 =	9 – 3 =
5 – 2 =	9 – 2 =
6 – 4 =	7 – 2 =
3 – 1 =	6 – 3 =
6 – 5 =	9 – 6 =
7 – 4 =	6 – 5 =
3 – 2 =	9 – 4 =
7 – 5 =	5 – 3 =
8 – 3 =	8 – 6 =
7 – 1 =	5 – 1 =
9 – 3 =	9 – 8 =

Appendix 14: Symbolic Approximate Arithmetic

	Symbolic	Sarah	John
1	Sarah has 6 candies …she gets 6 more John has 15. Who has more?		
2	S. has 15 and gets 25 more. John has 50. Who has more?		
3	S. has 8 and gets 6 more. J. has 21. Who has more?		
4	S. has 9 and gets 12 more. J. has 14. Who has more?		
5	S. has 7 and gets 9 more. J. has 20. Who has more?		
6	S. has 27 and gets 31 more. J. has 33. Who has more?		
7	S. has 10 and gets 11 more. J. has 12. Who has more?		
8	S. has 16 and gets 16 more. J. has 56. Who has more?		
9	S. has 11 and gets 12 more. J. has 13. Who has more?		
10	S. has 9 and gets 6 more. J. has 12. Who has more?		
11	S. has 5 and gets 5 more. J. has 15. Who has more?		
12	S. has 21 and gets 30 more. J. has 34. Who has more?		
13	S. has 25 and gets 20 more. J. has 36. Who has more?		
14	S. has 6 and gets 6 more. J. has 21. Who has more?		

15	S. has 25 and gets 20 more. J. has 30. Who has more?		
16	S. has 20 and gets 30 more. J. has 40. Who has more?		
17	S. has 30 and gets 26 more. J. has 32. Who has more?		
18	S. has 9 and gets 6 more. J. has 10. Who has more?		
19	S. has 16 and gets 17 more. J. has 58. Who has more?		
20	S. has 15 and gets 19. J. has 51. Who has more?		
21	S. has 20 and gets 16 more. J. has 45. Who has more?		
22	S. has 12 and gets 8 more. J. has 16. Who has more?		
23	S. has 6 and gets 7 more. J. has 23. Who has more?		
24	S. has 15 and gets 15 more. J. has 51. Who has more?		

Appendix 15: Nonsymbolic Approximate Arithmetic

	Non-Symbolic	Sarah	John
1	Sarah has that many marbles …she gets that many more John has that many marbles. Who has more?		
2	S. has that many marbles and gets that many more. J. has that many marbles.		
3	S. has that many marbles and gets that many more. J. has that many marbles.		
4	S. has that many marbles and gets that many more. J. has that many marbles.		
5	S. has that many marbles and gets that many more. J. has that many marbles.		
6	S. has that many marbles and gets that many more. J. has that many marbles.		
7	S. has that many marbles and gets that many more. J. has that many marbles.		
8	S. has that many marbles and gets that many more. J. has that many marbles.		
9	S. has that many marbles and gets that many more. J. has that many marbles.		
10	S. has that many marbles and gets that many more. J. has that many marbles.		
11	S. has that many marbles and gets that many more. J. has that many marbles.		
12	S. has that many marbles and gets that many more. J. has that many marbles.		

13	S. has that many marbles and gets that many more. J. has that many marbles.		
14	S. has that many marbles and gets that many more. J. has that many marbles.		
15	S. has that many marbles and gets that many more. J. has that many marbles.		
16	S. has that many marbles and gets that many more. J. has that many marbles.		
17	S. has that many marbles and gets that many more. J. has that many marbles.		
18	S. has that many marbles and gets that many more. J. has that many marbles.		
19	S. has that many marbles and gets that many more. J. has that many marbles.		
20	S. has that many marbles and gets that many more. J. has that many marbles.		
21	S. has that many marbles and gets that many more. J. has that many marbles.		
22	S. has that many marbles and gets that many more. J. has that many marbles.		
23	S. has that many marbles and gets that many more. J. has that many marbles.		
24	S. has that many marbles and gets that many more. J. has that many marbles.		

Appendix 16: Head-Toes-Shoulders-Knees Task

Part 1

Training		Incorrect	Self-Correct	Correct
A1	What would you do if I say "touch your toes?" a) verbal response B) behavioural response	0	1	2
A2	What would you do if I say "touch your toes?" a) verbal response B) behavioural response	0	1	2

Practice		0	1	2
B1	Touch your head	0	1	2
B2	Touch your toes	0	1	2
B3	Touch your head	0	1	2
B4	Touch your toes	0	1	2

Test		0	1	2
1	Touch your head	0	1	2
2	Touch your toes	0	1	2
3	Touch your toes	0	1	2
4	Touch your head	0	1	2
5	Touch your toes	0	1	2
6	Touch your head	0	1	2
7	Touch your head	0	1	2
8	Touch your toes	0	1	2
9	Touch your head	0	1	2
10	Touch your toes	0	1	2

Part 2

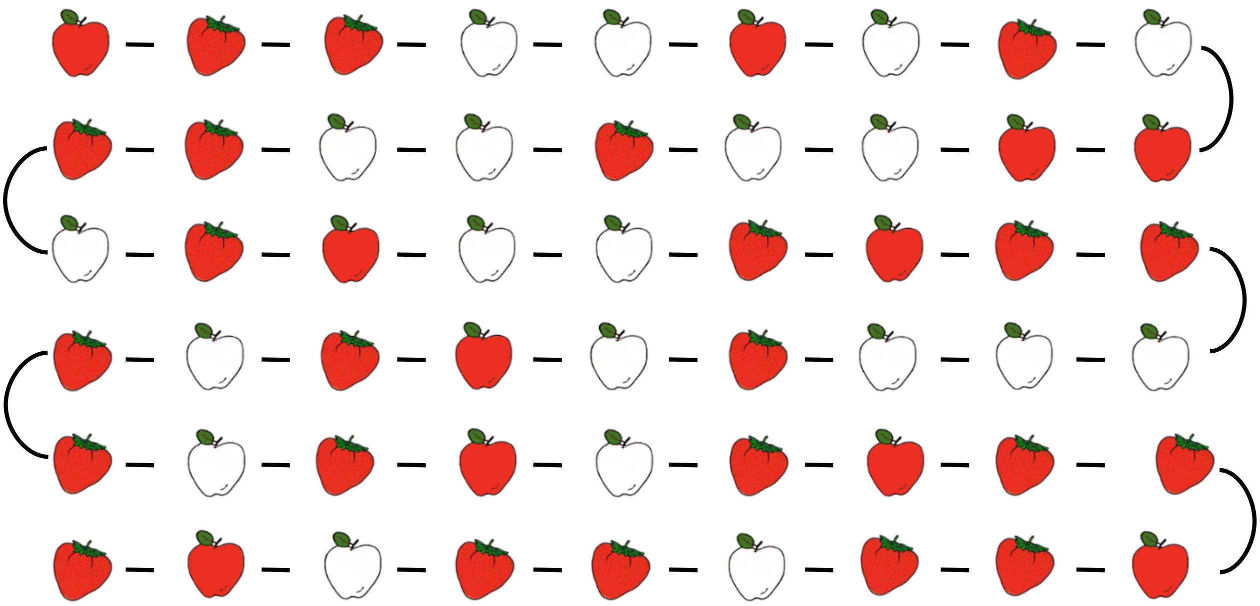
Training		Incorrect	Self-Correct	Correct
C1	What would you do if I say "touch your knees?" a) verbal response B) behavioural response	0	1	2
C2	What would you do if I say "touch your shoulders?" a) verbal response B) behavioural response	0	1	2

Practice		0	1	2
D1	Touch your knees	0	1	2
D2	Touch your shoulders	0	1	2
D3	Touch your knees	0	1	2
D4	Touch your shoulders	0	1	2

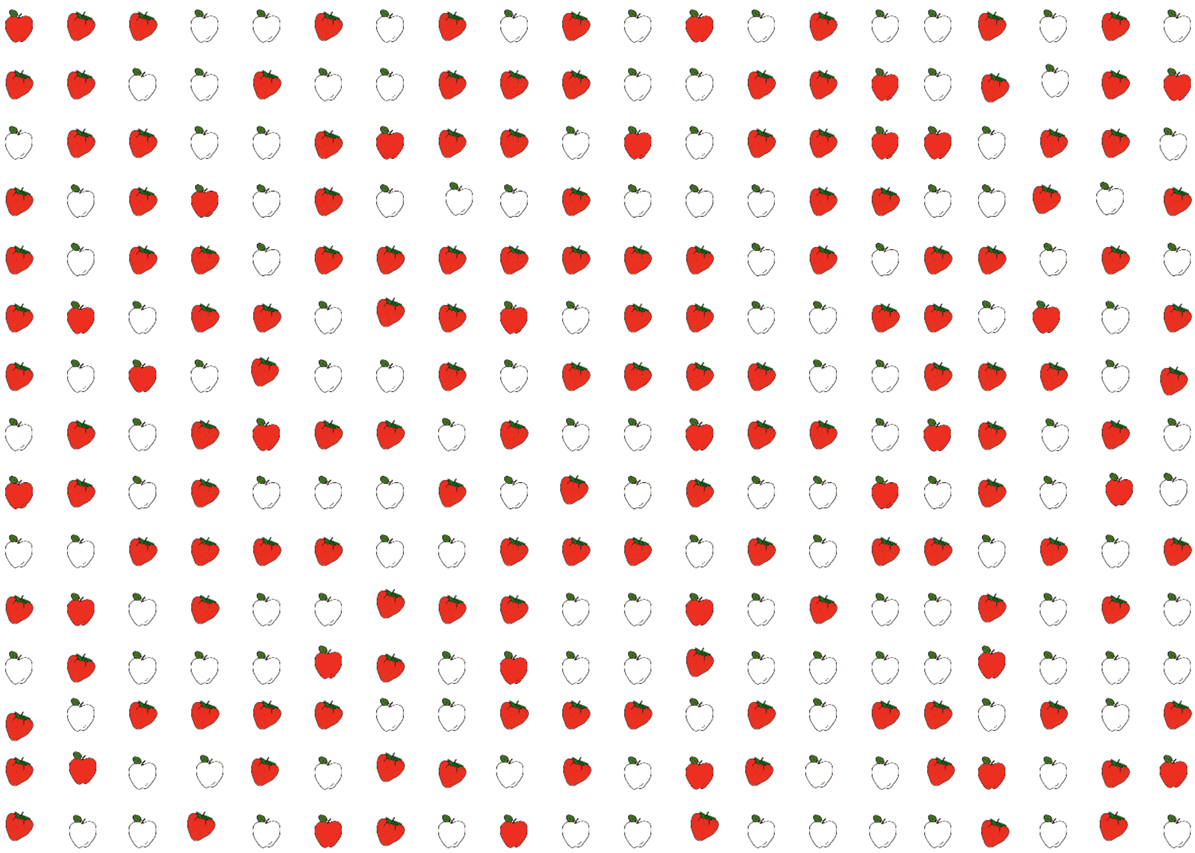
Test		0	1	2
11	Touch your head	0	1	2
12	Touch your toes	0	1	2
13	Touch your knees	0	1	2
14	Touch your toes	0	1	2
15	Touch your shoulders	0	1	2
16	Touch your head	0	1	2
17	Touch your knees	0	1	2
18	Touch your knees	0	1	2
19	Touch your shoulders	0	1	2
20	Touch your toes	0	1	2

Appendix 17: Visual Search Task

Easy Part Time 1



Hard Part Time 1



Appendix 18: Calculation of d' (d prime)

Variables recorded: Hits = Number of correctly crossed out red apples (out of 17)

Missed = Number of missed red apples

Strawberries = Number of wrongly crossed out red strawberries (out of 37)

White apples = Number of wrongly crossed out white apples (out of 36)

Correct-Rejects = Number of correctly rejected distractors (out of 73)

False-Alarm = Strawberries + White apples

Hit-Rate = Hits / (Hits + Missed)

False-Alarm-Rate = False-Alarm / (False-Alarm + Correct-Rejects)

d' = z(Hit-Rate) – z(False-Alarm-Rate)

Online d'- calculator: http://memory.psych.mun.ca/models/dprime/

Appendix 19: Booklet Order of Magnitude Comparison at Times 3, 4 and 5

Booklet 1

Nonsymbolic fixed size far

Nonsymbolic surface-area matched ratio 5:6

Symbolic close

Nonsymbolic fixed size ratio 3:4

Nonsymbolic surface-area matched ratio 2:3

Booklet 2

Symbolic far

Nonsymbolic fixed size ratio 2:3

Nonsymbolic surface-area matched far

Nonsymbolic fixed size ratio 5:6

Nonsymbolic fixed size close

Nonsymbolic surface-area matched ratio 3:4

Appendix 20: Standardized coefficients of Structural Equation Modelling of Concurrent Associations of arithmetic at Times 1, 2, 3, 4 and 5

	Time 1			Time 2			Time 3			Time 4			Time 5	
	Estimate	p-value		Estimate	p-value		Estimate	p-value		Estimate	p-value		Estimate	p-value
Nonverbal IQ with			Math-Language with			Counting with			Transcoding with			Executive with		
Language	.451	.01	Transcoding	.554	<.001	Transcoding	.880	<.001	Executive Function	.528	.015	Literacy	.469	<.001
Math-Language	.439	.011	Counting	.342	.046	Digit Magnitude	.665	<.001	Magnitude	.658	<.001	Transcoding	.521	<.001
Transcoding	.409	.073	Digit Magnitude	.582	<.001	NS Magnitude	.587	<.001				Magnitude	.693	<.001
Counting	.229	.308	NS Magnitude	.637	<.001									
Digit Magnitude	.437	.033												
NS Magnitude	.632	<.001												
Language with			Transcoding with			Transcoding with			Executive with			Literacy with		
Math-Language	.621	<.001	Counting	.774	<.001	Digit Magnitude	.794	<.001	Magnitude	.831	<.001	Transcoding	.724	<.001
Transcoding	.529	<.001	Digit Magnitude	.735	<.001	NS Magnitude	.636	<.001				Magnitude	.545	<.001
Counting	.290	.058	NS Magnitude	.495	<.001									
Digit Magnitude	.605	<.001												
NS Magnitude	.469	<.001												
Math-Language with			Counting with			Digit with						Transcoding with		
Transcoding	.346	.008	Digit Magnitude	.622	<.001	NS Magnitude	.869	<.001				Magnitude	.493	<.001
Counting	.212	.208	NS Magnitude	.549	<.001									
Digit Magnitude	.511	<.001												
NS Magnitude	.376	.007												
Transcoding with			Digit with											
Counting	.666	<.001	NS Magnitude	.603	<.001									
Digit Magnitude	.672	<.001												
NS Magnitude	.415	<.001												
Counting with														
Digit Magnitude	.406	.011												
NS Magnitude	.332	.024												
Digit with														
NS Magnitude	.556	<.001												

Appendix 21: Standardized coefficients SEM of Longitudinal Prediction of arithmetic at Times 2, 3, 4 and 5 by Time 1 base model

	Time 2	
	Estimate	p-value
Nonverbal IQ T1 with		
Language T1	.238	.297
Math-Language T1	.247	.302
Transcoding T1	.403	.045
Counting T1	.185	.470
Digit Magnitude T1	.218	.374
NS Magnitude T1	.583	.006
Language T1 with		
Math-Language T1	.585	<.001
Transcoding T1	.510	<.001
Counting T1	.282	.065
Digit Magnitude T1	.602	<.001
NS Magnitude T1	.457	
Math-Lang. T1 with		
Transcoding T1	.322	.011
Counting T1	.216	.189
Digit Magnitude T1	.506	<.001
NS Magnitude T1	.378	.004
Transcoding T1 with		
Counting T1	.670	<.001
Digit Magnitude T1	.638	<.001
NS Magnitude T1	.412	<.001
Counting T1 with		
Digit Magnitude T1	.388	.017
NS Magnitude T1	.317	.029
Digit T1 with		
NS Magnitude T1	.547	<.001

	Time 3	
	Estimate	p-value
Nonverbal IQ T1 with		
Language T1	.240	.292
Math-Language T1	.199	.418
Transcoding T1	.395	.046
Counting T1	.181	.479
Digit Magnitude T1	.218	.374
NS Magnitude T1	.582	.006
Language T1 with		
Math-Language T1	.620	<.001
Transcoding T1	.511	<.001
Counting T1	.277	.072
Digit Magnitude T1	.598	<.001
NS Magnitude T1	.458	<.001
Math-Lang. T1 with		
Transcoding T1	.325	.009
Counting T1	.205	.222
Digit Magnitude T1	.512	<.001
NS Magnitude T1	.375	.005
Transcoding T1 with		
Counting T1	.669	<.001
Digit Magnitude T1	.628	<.001
NS Magnitude T1	.411	<.001
Counting T1 with		
Digit Magnitude T1	.386	.018
NS Magnitude T1	.314	.030
Digit T1 with		
NS Magnitude T1	.546	<.001

	Time 4	
	Estimate	p-value
Nonverbal IQ T1 with		
Language T1	.234	.307
Math-Language T1	.195	.430
Transcoding T1	.411	.039
Counting T1	.172	.504
Digit Magnitude T1	.205	.404
NS Magnitude T1	.578	.007
Language T1 with		
Math-Language T1	.618	<.001
Transcoding T1	.491	<.001
Counting T1	.268	.083
Digit Magnitude T1	.594	<.001
NS Magnitude T1	.453	<.001
Math-Lang. T1 with		
Transcoding T1	.323	.009
Counting T1	.199	.239
Digit Magnitude T1	.511	<.001
NS Magnitude T1	.372	.005
Transcoding T1 with		
Counting T1	.674	<.001
Digit Magnitude T1	.634	<.001
NS Magnitude T1	.400	<.001
Counting T1 with		
Digit Magnitude T1	.371	.025
NS Magnitude T1	.306	.035
Digit T1 with		
NS Magnitude T1	.540	<.001

	Time 5	
	Estimate	p-value
Nonverbal IQ T1 with		
Language T1	.249	.274
Math-Language T1	.206	.400
Transcoding T1	.416	.041
Counting T1	.195	.444
Digit Magnitude T1	.223	.362
NS Magnitude T1	.522	.009
Language T1 with		
Math-Language T1	.624	<.001
Transcoding T1	.520	<.001
Counting T1	.294	.054
Digit Magnitude T1	.608	<.001
NS Magnitude T1	.411	<.001
Math-Lang. T1 with		
Transcoding T1	.343	.008
Counting T1	.216	.197
Digit Magnitude T1	.520	<.001
NS Magnitude T1	.360	.002
Transcoding T1 with		
Counting T1	.679	<.001
Digit Magnitude T1	.656	<.001
NS Magnitude T1	.372	<.001
Counting T1 with		
Digit Magnitude T1	.394	.015
NS Magnitude T1	.277	.046
Digit T1 with		
NS Magnitude T1	.504	<.001

Appendix 22: Standardized coefficients of SEM of Number Wizards' Concurrent Associations at Time 1 and Longitudinal Prediction of arithmetic at Time 5 by Time 1 base model

	Concurrent Time 1			Longitudinal Time 5	
	Estimate	p-value		Estimate	p-value
Nonverbal IQ T1 with			Nonverbal IQ T1 with		
Language T1	.513	.090	Language T1	.497	.100
Math-Language T1	-.067	.843	Math-Language T1	.007	.983
Transcoding T1	.511	.049	Transcoding T1	.564	.028
Counting T1	.098	.736	Counting T1	.098	.736
Magnitude T1	.499	.092	Magnitude T1	.371	.233
Language T1 with			Language T1 with		
Math-Language T1	.708	<.001	Math-Language T1	.740	<.001
Transcoding T1	.408	.027	Transcoding T1	.414	.023
Counting T1	.142	.505	Counting T1	.133	.525
Magnitude T1	.669	<.001	Magnitude T1	.633	.001
Math-Lang. T1 with			Math-Lang. T1 with		
Transcoding T1	.254	.236	Transcoding T1	.255	.234
Counting T1	.156	.486	Counting T1	.185	.417
Magnitude T1	.615	.001	Magnitude T1	.628	.003
Transcoding T1 with			Transcoding T1 with		
Counting T1	.641	<.001	Counting T1	.640	<.001
Magnitude T1	.597	<.001	Magnitude T1	.578	.002
Counting T1 with			Counting T1 with		
Magnitude T1	.374	.051	Magnitude T1	.300	.152

Appendix 23: Standardized coefficients SEM of Symbolic Approximate Arithmetic at Times 3, 4 and 5 by Time 1 base model

	Time 3			Time 4			Time 5	
	Estimate	p-value		Estimate	p-value		Estimate	p-value
Nonverbal IQ T1 with			Nonverbal IQ T1 with			Nonverbal IQ T1 with		
Language T1	.242	.289	Language T1	.244	.285	Language T1	.245	.283
Math-Language T1	.206	.403	Math-Language T1	.206	.401	Math-Language T1	.207	.398
Transcoding T1	.394	.057	Transcoding T1	.384	.064	Transcoding T1	.414	.044
Counting T1	.192	.453	Counting T1	.192	.452	Counting T1	.194	.448
Digit Magnitude T1	.275	.256	Digit Magnitude T1	.273	.259	Digit Magnitude T1	.293	.226
NS Magnitude T1	.585	.006	NS Magnitude T1	.585	.006	NS Magnitude T1	.586	.006
Language T1 with			Language T1 with			Language T1 with		
Math-Language T1	.623	<.001	Math-Language T1	.624	<.001	Math-Language T1	.624	<.001
Transcoding T1	.528	<.001	Transcoding T1	.509	<.001	Transcoding T1	.525	<.001
Counting T1	.295	.053	Counting T1	.296	.052	Counting T1	.298	.050
Digit Magnitude T1	.628	<.001	Digit Magnitude T1	.620	<.001	Digit Magnitude T1	.629	<.001
NS Magnitude T1	.462	<.001	NS Magnitude T1	.464	<.001	NS Magnitude T1	.464	<.001
Math-Lang. T1 with			Math-Lang. T1 with			Math-Lang. T1 with		
Transcoding T1	.342	.009	Transcoding T1	.346	.008	Transcoding T1	.344	.008
Counting T1	.215	.200	Counting T1	.216	.198	Counting T1	.217	.195
Digit Magnitude T1	.446	.002	Digit Magnitude T1	.449	.002	Digit Magnitude T1	.446	.002
NS Magnitude T1	.382	.004	NS Magnitude T1	.382	.004	NS Magnitude T1	.383	.004
Transcoding T1 with			Transcoding T1 with			Transcoding T1 with		
Counting T1	.667	<.001	Counting T1	.670	<.001	Counting T1	674	<.001
Digit Magnitude T1	.709	<.001	Digit Magnitude T1	.715	<.001	Digit Magnitude T1	.711	<.001
NS Magnitude T1	.414	<.001	NS Magnitude T1	.413	<.001	NS Magnitude T1	.418	<.001
Counting T1 with			Counting T1 with			Counting T1 with		
Digit Magnitude T1	.405	.009	Digit Magnitude T1	.415	.007	Digit Magnitude T1	.414	.007
NS Magnitude T1	.324	.025	NS Magnitude T1	.325	.025	NS Magnitude T1	.326	.024
Digit T1 with			Digit T1 with			Digit T1 with		
NS Magnitude T1	.505	<.001	NS Magnitude T1	.507	<.001	NS Magnitude T1	.513	<.001

Appendix 24: Standardized coefficients SEM of Nonsymbolic Approximate Arithmetic at Times 3, 4 and 5 by Time 1 base model

	Time 3	
	Estimate	p-value
Nonverbal IQ T1 with		
Language T1	.241	.293
Math-Language T1	.204	.407
Transcoding T1	.392	.057
Counting T1	.190	.458
Digit Magnitude T1	.247	.302
NS Magnitude T1	.588	.006
Language T1 with		
Math-Language T1	.621	<.001
Transcoding T1	.524	<.001
Counting T1	.291	.057
Digit Magnitude T1	.590	<.001
NS Magnitude T1	.463	<.001
Math-Lang. T1 with		
Transcoding T1	.337	.009
Counting T1	.212	.207
Digit Magnitude T1	.493	<.001
NS Magnitude T1	.379	.005
Transcoding T1 with		
Counting T1	.667	<.001
Digit Magnitude T1	.665	<.001
NS Magnitude T1	.414	<.001
Counting T1 with		
Digit Magnitude T1	.441	.004
NS Magnitude T1	.325	.026
Digit T1 with		
NS Magnitude T1	.512	<.001

	Time 4	
	Estimate	p-value
Nonverbal IQ T1 with		
Language T1	-.022	.926
Math-Language T1	.455	.043
Transcoding T1	.402	.060
Counting T1	.312	.174
Digit Magnitude T1	.104	.699
NS Magnitude T1	.834	<.001
Language T1 with		
Math-Language T1	.606	<.001
Transcoding T1	.519	<.001
Counting T1	.296	.050
Digit Magnitude T1	.601	<.001
NS Magnitude T1	.446	<.001
Math-Lang. T1 with		
Transcoding T1	.333	.011
Counting T1	.208	.216
Digit Magnitude T1	.487	.001
NS Magnitude T1	.363	.010
Transcoding T1 with		
Counting T1	663.	<.001
Digit Magnitude T1	.662	<.001
NS Magnitude T1	.410	<.001
Counting T1 with		
Digit Magnitude T1	.433	.005
NS Magnitude T1	.335	.024
Digit T1 with		
NS Magnitude T1	.503	<.001

	Time 5	
	Estimate	p-value
Nonverbal IQ T1 with		
Language T1	.246	.281
Math-Language T1	.208	.397
Transcoding T1	.393	.057
Counting T1	.195	.445
Digit Magnitude T1	.258	.282
NS Magnitude T1	.590	.006
Language T1 with		
Math-Language T1	.624	<.001
Transcoding T1	.522	<.001
Counting T1	.299	.050
Digit Magnitude T1	.592	<.001
NS Magnitude T1	.468	<.001
Math-Lang. T1 with		
Transcoding T1	.342	.009
Counting T1	.219	.191
Digit Magnitude T1	.496	<.001
NS Magnitude T1	.383	.005
Transcoding T1 with		
Counting T1	.675	<.001
Digit Magnitude T1	.681	<.001
NS Magnitude T1	.414	<.001
Counting T1 with		
Digit Magnitude T1	.460	.003
NS Magnitude T1	.332	.022
Digit T1 with		
NS Magnitude T1	.513	<.001

Appendix 25: Standardized coefficients SEM of Longitudinal Prediction of arithmetic at Time 5 by Time 1 base model and number line estimation

	Time 5	
	Estimate	p-value
Arithmetic Time 1 with		
Transcoding T1	.623	<.001
Counting T1	.405	.002
Symbolic Magnitude T1	.559	<.001
Nonsymbolic Magnitude T1	.513	<.001
Number Line 0-10: small numbers	-.062	.648
Number line 0-10: large numbers	.124	.483
Transcoding T1 with		
Counting T1	.684	<.001
Symbolic Magnitude T1	.634	<.001
Nonsymbolic Magnitude T1	.431	<.001
Number Line 0-10: small numbers	-.179	.173
Number line 0-10: large numbers	.222	.123
Counting T1 with		
Symbolic Magnitude T1	.365	.031
Nonsymbolic Magnitude T1	.344	.021
Number Line 0-10: small numbers	-.337	.426
Number line 0-10: large numbers	.195	.334
Symbolic Magnitude T1 with		
Nonsymbolic Magnitude T1	.535	<.001
Number Line 0-10: small numbers	-.337	.020
Number line 0-10: large numbers	-.095	.583
Nonsymbolic Magnitude T1 with		
Number Line 0-10: small numbers	-.142	.321
Number line 0-10: large numbers	-.035	.818
Number line 0-10: small numbers T1 with		
Number line 0-10: large numbers T1	-.092	.622

List of References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (DSM-IV)*. Washington DC: American Psychiatric Association.

Ansari, D., Donlan, C., Thomas, M., Ewing, S., Peen, T., Karmiloff-Smith, A. (2003). What makes counting count? Verbal and visuo-spatial contributions to typical and atypical number development. *Journal of Experimental Child Psychology, 85,* 50-62.

Aunola, K., Leskinen, E., Lerkkanen, M.-K., and Nurmi, J. E. (2004). Developmental dynamics of math performance from pre-school to Grade 2. *Journal of Educational Psychology, 96,* 699–713.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4(11),* 417-423.

Baldo, J. and Dronkers, N. (2007). Neural correlates of arithmetic and language comprehension: A common substrate? *Neuropsychologia, 45,* 229-235.

Barth, H., Kanwisher, N., and Spelke, E. (2003). The construction of large number representations in adults. *Cognition, 86,* 201-221.

Barth, H., LaMont, K., Lipton, J., and Spelke, E. (2005). Abstract Number and Arithmetic in Pre-school Children. *PNAS, 102,* 14116-14121.

Barth, H., LaMont, K., Lipton, J., Dehaene, S., Kanwisher, N., and Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition, 98,* 199-222.

Berg, D. H. (2008). Working memory and arithmetic calculation in children: The contributory roles of processing speed, short term memory, and reading. *Journal of Experimental Child Psychology, 99,* 288-308.

Bishop, D. (2003). *Test for Reception of Grammar-Version 2 (TROG-2).* San Antonio, TX: The Psychological Corporation.

Boonen, A. J. H., Kolkman, M. E., and Kroesbergen, E. H. (2011). The relation between teachers’ math talk and the acquisition of number sense within kindergarten classrooms. *Journal of School Psychology, 49,* 281-299.

Booth, J. L., and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 42,* 189-201.

Booth, J. L., and Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79,* 1016-1031.

Brigstocke, S., Moll, K., and Hulme, C. (2016). *TOBANS: Test of Basic Arithmetic and Numeracy Skills*. Oxford, UK: Oxford University Press.

Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., and Grimm, K. J. (2009). The contributions of 'hot' and 'cool' executive function to children's academic achievement, learning related behaviours, and engagement in kindergarten. *Early Childhood Research Quarterly, 24,* 713-721.

Browne, M. W., and Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research, 23,* 230-258.

Bull, R., Espy, K. A., and Wiebe, S. A. (2008). Short-Term Memory, Working Memory, and Executive Functioning in Pre-schoolers: Longitudinal Predictors of Mathematical Achievement at Age 7 Years. *Journal of Developmental Neuropsychology, 33,* 205-228.

Bull, R., Epsy, K. A., Wiebe, S. A., Sheffield, T. D., and Nelson, J. M. (2011). Using confirmatory factor analysis to understand executive control in pre-school children: sources of variation in emergent mathematic achievement. *Developmental Science, 14,* 679-692.

Bull, R., and Scerif, G. (2001). Executive functioning as a predictor of children's mathematical ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19,* 273-293.

Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry, 46,* 3-18.

Byrne, B. M. (2012). *Structural equation modeling with MPlus: Basic concepts, applications and programming.* London: Routledge.

Cameron Ponitz, C. E., McClelland, M. M., Jewkes, A. M., Connor, C., Farris, C. L., and Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioural regulation in early childhood. *Early Childhood Research Quarterly, 23,* 141-158.

Campbell, J. I. D. (1995). Mechanisms of Simple Addition and Multiplication: A Modified Network-interference Theory and Simulation. *Mathematical Cognition, 1,* 121-164.

Campbell, J. I. D., and Clark, J. M. (1988). An encoding-complex view of cognitive number processing: Comment on McCloskey, Sokol & Goodman (1986). *Journal of Experimental Psychology: General, 117,* 204-214.

Caravolas, M., Lervåg, M., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., Salas, N., Schöffelová, M., Defior, S., Mikulajová, M., Seidlová-Málková, G., and Hulme, C. (2012). Common Patterns of Prediction of Literacy Development in Different Alphabetic Orthographies. *Psychological Science, 23,* 678-686.

Chiappe, P., Siegel, L. S., and Hasher L. (2000). Working memory, inhibitory control, and reading disability. *Memory & Cognition, 28,* 8-17.

Clark, E. V. (1973). Non-linguistic strategies and the acquisition of word meanings. *Cognition, 2*, 161-182.

Clements, D. H. (1999). Subitizing: What is it? Why teach it? *Teaching Children Mathematics, 5,* 400-405.

Cowan, R., Donlan, C., Newton, E. J., and Lloyd D. (2005). Number skills and knowledge in children with specific language impairment. *Journal of Educational Psychology. 97*, 732-744.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44,* 1-42.

Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics.* London: Allen Lane, The Penguin Press.

Dehaene, S., Dupoux, E., and Mehler, J. (1990). Is numerical comparison a? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception Performance, 16,* 626-641/

Dehaene, S., and Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition, 43,* 1-29.

Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20,* 487-506.

de Jong .(1999). Hierarchical regression analysis in structural equation modelling. *Structural Equation Modeling: A Multidisciplinary Journal, 6,* 198-211.

De Smedt, B., Noël, M. P., Gilmore, D., and Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education, 2,* 48–55

Desoete, A., Ceulemans, A., De Weerdt, F., and Pieters, S. (2012). Can we predict mathematical learning disabilities from symbolic and non-symbolic comparison tasks in kindergarten? Findings from a longitudinal study. *British Journal of Educational Psychology, 82,* 64–81.

Desoete, A., and Grégoire, J. (2006). Numerical competence in young children and in children with mathematical learning abilities. *Learning and Individual Differences, 16,* 351-367.

Donaldson, M., and Balfour, G. (1968). A study of language comprehension in children. *British Journal of Psychology, 59*, 461-471.

Donlan, C., Bishop, D. V. M., and Hitch, G. J. (1998). Magnitude comparison by children with specific language impairments: Evidence of unimpaired symbolic processing. *Journal of Language and Communication Disorders, 33*, 149-160.

Donlan, C., Cowan, R. Newton, E. J., and Lloyd, D. (2007). The role of mathematical development: Evidence from children with specific language impairments. *Cognition, 103,* 23-33.

Donlan, C., and Gourlay, S. (1999). The importance of non-verbal skills in the acquisition of place-value knowledge: Evidence from normally developing and language-impaired children. *British Journal of Developmental Psychology*, *17*, 1-19.

Dowker, A. (2005). *Individual differences in arithmetic: Implications for psychology, neuroscience, and education*. New York: Psychology Press.

Dunn, L. M., Dunn, D. M., and Styles, B. (2010). *British Picture Vocabulary Scale (3rd ed.).* London, England: GL Assessment.

Durand, M., Hulme, C., Larkin, R., and Snowling, M. (2005). The cognitive foundations of reading and arithmetic skills in 7-to 10-year-olds. *Journal of Experimental Child Psychology, 91*, 113-136.

Edelman, B., Abdi, H., and Valentin, D. (1996). Multiplication Number Facts: Modeling Human Performance with Connectionist Networks. *Psychologica Belgica, 36,* 31-63.

Fazio, B. B. (1994). The counting abilities of children with specific language impairment: A comparison of oral and gestural tasks. *Journal of Speech and Hearing Research, 37,* 358-368.

Fazio, B. B. (1996). Mathematical abilities of children with specific language impairment: A two-year follow-up. *Journal of Speech and Hearing Research, 39*, 839-849.

Friso-van den Bos, I., Kroesbergen, E. H., Van Luit, J. E. H., Xenidou-Dervou, I., Jonkman, L. M., Van der Schoot, M., and Van Lieshout, E. C. D. M. (2015). Longitudinal development of number line estimation and mathematics performance in primary school children. *Journal of Experimental Child Psychology, 134,* 12-29.

Fuhs, M. W., and McNeil, N. M. (2013). ANS acuity and mathematics ability in pre-schoolers from low-income homes: contributions of inhibitory control. *Developmental Science, 16,* 136-148.

Gathercole, S. E., and Pickering, S. J. (2000).Working memory deficits in children with low achievements in the national curriculum at seven years of age. *British Journal of Educational Psychology, 70,* 177-194.

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from National Curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology, 40,* 1-16.

Geary, D., Hoard, M. K., Byrd-Craven, J., Nugent, L., and Numtee, C. (2007). Cognitive Mechanisms Underlying Achievement Deficits in Children With Mathematical Learning Disability. *Child Development, 78*, 1343-1359.

Geary, D., Hoard, M. K., Nugent, L., and Bailey, D. H. (2012). Mathematical Cognition Deficits in Children With Learning Disabilities and Persistent Low Achievement: A Five-Year Prospective Study. *Journal of Educational Psychology, 104,* 206-223.

Geary, D. C., and Wiley, J. G. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in young and elderly adults. *Psychology & Ageing, 6,* 474-483.

Gelman, R., and Gallistel, C. R. (1978). *The child's understanding of number.* Cambridge, MA: Harvard Press.

Gilmore, K., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., Simms, V., and Inglis, M. (2013). Individual Differences in Inhibitory Control, Not Non-verbal Number Acuity, Correlate with Mathematics Achievement. *PLoS ONE, 8,* 1-9.

Gilmore, C., Keeble, S., Richardson, S., and Cragg, L. (2014). The role of cognitive inhibition in different components of arithmetic. *ZDM Mathematics Education, 47,* 771-782.

Gilmore, C., McCarthy, S. E., and Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature, 447*, 589-591.

Gilmore, C., McCarthy, S. E., and Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition, 115,* 394-406.

Göbel, S. M., Watson, S. E., Lervåg, A., and Hulme, C. (2014). Children's Arithmetic Development: It Is Number Knowledge, Not the Approximate Number Sense, That Counts. *Psychological Science, 25,* 789 –798.

Halberda, J., Mazzocco, M. M. M., and Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*, 665-668.

Holloway, I. D., and Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of experimental Child Psychology, 103,* 17-29.

References

Holmes, J., Adams, J. W., and Hamilton, C. J. (2006). The relationship between visuospatial sketchpad capacity and children's mathematical skills. *European Journal of Cognitive Psychology, 20,* 272-289.

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6,* 1-55.

Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F. J., and Snowling, M. J. (2012). The causal role of phoneme awareness and letter-sound knowledge in learning to read combining intervention studies with mediation analyses. *Psychological Science, 23,* 572-577.

Hurford, J. (1987). *Language and Number.* Basil Blackwell: New York.

IBM (2011). IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.Iuculano, T., Tang, J, Hall, C. W. B., and Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Developmental Science, 11,* 669-680.

Jarvis, H. L., and Gathercole, S. E. (2003).Verbal and non- verbal working memory and achievements on national curriculum tests at 11 and 14 years of age. *Educational and Child Psychology, 20,* 123-140.

Jordan, N. C., Huttenlocher, J., and Levine, S. C. (1994). Assessing early arithmetic abilities: effects of verbal and nonverbal response types on the calculation performance of middle- and low-income children. *Learning and Individual Differences. 6,* 413-432.

Jordan, N. C., Kaplan, D, and Hanich, L. B. (2002). Achievement growth in children with learning difficulties in mathematics: Findings of a two-year longitudinal study. *Journal of Educational Psychology, 94,* 586-597.

Jordan, N. C, Kaplan, D., Ramineni, and Locuniak, M. N. (2009). Early Math Matters: Kindergarten Number Competence and Later Mathematics Outcomes. *Journal of Developmental Psychology, 45,* 850-867.

Krajewwski, K., and Scheider, W. (2009). Exploring the impact of phonological awareness, visual-spatial working memory, and pre-school quantaty-number competencies on mathematics achievement in elementary school: Findings from a 3-year longitudinal study. *Journal of Experimental Child Psychology, 103*, 516-531.

Kleemans, T., Segers, E., and Verhoeven, L. (2011). Precursors to numeracy in kindergartners with specific language impairment. *Research in Developmental Disabilities, 32,* 2901-2908.

Kleemans, T., Segers, E., and Verhoeven, L. (2012). Naming speed as a clinical marker in predicting basic calculation skills in children with specific language impairment. *Research in Developmental Disabilities, 33,* 882-889.

Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Hedges, L. V., and Vasilyeva, M. (2006). Pre-school Children's Mathematical Knowledge: The Effect of Teacher "Math Talk". *Developmental Psychology, 42*, 59-69.

Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling*, New York, NY: Guilford Press.

Kolkman, M. E., Kroesbergen, E. H., and Lesemann, P. P. (2012). Early numerical development and the role of non-symbolic and symbolic skills. *Learning and Instruction, 25,* 95–103

Korkman, M., Kirk, U., and Kemp, S. (1998). *NEPSY- A Developmental Neuropsychological Assessment,* Psychological Corporation.

LeFevre, J., Fast, L., Skwarchuk, S., Smith-Chant, B. L., Bisanz, J., Kamawar, D., and Penner-Wilger, M. (2010). Pathways to Mathematics: Longitudinal Predictors of Performance. *Child Development, 81*, 1753–1767.

LeFevre, J., Lira, C. J., Sowinski, C., Cankaya, O., Kamawar, D., and Skwarchuk, S. (2013). Charting the role of the number line in mathematical development. *Frontiers in Psychology, 4,* 1-9.

LeFevre, J., Sadesky, G. S., and Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem-size effect in adults. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22,* 216-230.

Lewis, C., Hitch, G. J., Walker, P. (1994). The Prevalence of Specific Arithmetic Difficulties and Specific Reading Difficulties in 9-to 10-year-old Boys and Girls. *Journal of Child Psychology and Psychiatry, 35,* 283-292.

Libertus, M. E., Feigenson, L., and Halberda, J. (2011). Pre-school acuity of the approximate number system correlates with school math ability. *Developmental Science, 14,* 1292–1300.

Libertus, M. E., Feigenson, L., and Halberda, J. (2013). Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. *Journal of Experimental Child Psychology, 116* 829-838.

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., and Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1-6. *Developmental Science, 17,* 714-726.

Matejko, A. A., and Ansari, D. (2016). Trajectories of Symbolic and Nonsymbolic Magnitude Processing in the First Year of Formal Schooling. *PLos ONE, 11(3),* e0149863.

Mazzocco, M. M. Feigenson, L., and Halberda, J. (2011). Pre-schoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE, 6(9),* e23749.

McClelland, M. M., Cameron, C. E., Connor, C., Farris, C. L., Jewkes, A. M., and Morrison, F. J. (2007). Links Between Behavioral Regulation and Pre-schoolers' Literacy, Vocabulary, and Math Skills. *Developmental Psychology, 43,* 947-959.

McCloskey, M., and Lindemann, M. A. (1992). MATHNET: Preliminary results from a distributed model of arithmetic fact retrieval. In J. I. D. Campbell, (Ed.), *The nature of and origins of mathematical skill* (pp. 365-409). Amsterdam: Elsevier Science.

McCloskey, M., Caramazza, A., and Basili, A. G. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition, 4,* 171-196.

McCloskey, M., Sokol, S. M., and Goodman, R. A. (1986). Cognitive processes in verbal-number production: Inferences from the performance of brain-damaged subjects. *Journal of Experimental Psychology: General, 115,* 307-330.

Miller, M. R., Müller, U., Giesbrecht, G. F., Carpendale, J. I., and Kerns, K. A. (2013). The contribution of executive function and social understanding to pre-schoolers' letter and math skills. *Cognitive Development, 28,* 331-349.

Mix, K. S., Prather, R W., Smith, L. B., and Stockton, J. D. (2014). Young Children's Interpretation of Multidigit Number Names: From Emerging Competence to Mastery. *Child Development, 85*, 1306-1319.

Möller, K., Pixner, S., Zuber, J., Kaufmann, L., and Nürk, H. C. (2011). Early place-value understanding as a precursor for later arithmetic performance – A longitudinal study on numerical development. *Research in Developmental Disabilities, 32,* 1837-1851.

Monette, S., Bigras, M., and Guay, M. C. (2011). The role of executive functions in school achievement at the end of Grad 1. *Journal of Experimental Child Psychology, 109,* 158-173.

Moyer, R. S., and Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature, 215,* 1519-1520.

Muldoon, K., Towse, J., Simms, V., Perra, O., and Menzies V. (2013). A longitudinal analysis of estimation, counting skills, and mathematical ability across the first school year. *Developmental Psychology, 49,* 250-257.

Mundy, E., and Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology, 103,* 490–502.

Muthén, I. K., and Muthén, B. O. (2013). *Mplus user's guide (7th ed.).* Los Angeles, CA: Author.

Neighbourhood Statistics. Offices for National Resources (2016). Retrieved from https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015.

Noël, M. P. (2009). Counting on working memory when learning to count on and to add: A pre-school study. *Developmental Psychology, 45(6),* 1660-1643.

Nunes, T., and Bryant, P. (1996). *Children doing mathematics*. Oxford, England: Blackwell.

Nunes, T., Bryant, P., Barros, R., and Sylva, K. (2012). The relative importance of two different mathematical abilities to mathematical achievement. *British Journal of Educational Psychology, 82,* 136-156.

Östergren, R., and Träff, U. (2013). Early number knowledge and cognitive ability affect early arithmetic ability. *Journal of Experimental Child Psychology, 115,* 405-421.

Passolunghi, M. C., and Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology, 88,* 348-367.

Piazza, M., and Dehaene, S. (2004). From number neurons to mental arithmetic: The cognitive neuroscience of number sense. In M. Gazzangia (Ed.). *The cognitive neurosciences* (3rd ed., pp. 865-875). Cambridge, MA: MIT Press.

Piazza, M. Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., Dehaene, S., and Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition, 116*, 33–41.

Praet, M., Titeca, D., Ceulemans, A., and Desoete, A. (2013). Language in the prediction of arithmetic in kindergarten and grade 1. *Learning and Individual Differences, 27,* 90-96.

Purpura, D. J., and Ganley, C. M. (2014). Working memory and language: Skill-specific or domain-general relations to mathematics? *Journal of Experimental Child Psychology, 122,* 104-121.

Ramini, G. B., Siegler, R. S., and Hitti, A. (2012). Taking it to the classroom: Number board games as a small group learning activity. *Journal of Educational Psychology, 104,* 661-672.

Rasmussen, C., and Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology, 91,* 137-157.

Raven, J. C, Court, J. H and Raven, J. C. (1993). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* Oxford: Oxford Psychologists Press.

References

Rugani, R., Vallortigara, G., Priftis, K., and Regolin, L. (2015). Number-space mapping in the newborn chick resembles humans' mental number line. *Science, 347,* 534-536.

Sacre, L., and Masterson, J. (2000). *Single word spelling test.* London: Nfer-Nelson.

Sasanguie, D., Göbel, S. M., Smets, K., and Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology, 114,* 418-431.

Saxe, G. B., Guberman, S. R., Gearhart, M., Gelman, R., Massey, C. M., and Rogoff, B. (1987). Social Processes in Early Number Development. *Monographs of the Society for Research in Child Development, 52,* (2, Whole No. 216).

Sekuler, R., and Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development, 48,* 630-633.

Sideridis, G., Simos, P., Papanicolaou, A., and Fletcher, J. (2014). Using Structural Equation Modeling to Assess Functional Connectivity in the Brain Power and Sample Size Considerations. *Educational and Psychological Measurement, 74,* 733-758.

Siegler, R. S. (2016). Magnitude knowledge: the common core of numerical development. *Developmental Science, 19,* 341-361.

Siegler, R. S., and Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75,* 428-444.

Siegler, R. S., and Braithwaite, D. W. (2017). Numerical Development. *Annual Review of Psychology, 68,* 187-213.

Siegler, R. S., and Lemaire, P. (1997). Older and Younger Adults' Strategy Choices in Multiplication: Testing Predictions of ASCM Using the Choice/No-Choice Method. *Journal of Experimental Psychology: General, 126,* 71-92.

Siegler, R. S., and Opfer, J. (2003). The Development of Numerical Estimation: Evidence for Multiple Representations of Numerical Quantity. *Psychological Science, 14,* 237-243.

Siegler, R. S., and Shipley, C. (1995). Variation, selection, and cognitive change. In T. Simon & G. Halford (Ed.), *Developing cognitive Competence: New approaches to process modelling* (pp. 31-76). Hillsdale, NJ: Erlbaum.

Siegler, R. ., Thompson, C. A., and Opfer, J. (2009). The Logarithmic-To-Linear Shift: One Learning Sequence, Many Tasks, Many Time Scales. *Mind, Brain, and Education*, *3,* 143-150.

Simmons, F. R., and Singleton, C. (2008). Do weak phonological representations impact on arithmetic development? A review of research into arithmetic and dyslexia. *Dyslexia*, *14*, 77-94.

Simmons, F. R., Willis, C., and Adams, A. (2012). Different components of working memory have different relationships with different mathematical skills. *Journal of Experimental Child Psychology*, *111,* 139-155.

Simms, V., Gilmore, C., Cragg, L., Marlow, N., Wolke, D., and Johnson, S. J. (2013). Mathematics difficulties in extremely preterm children: evidence of a specific deficit in basic mathematics processing. *Paediatric Research, 73,* 236-244.

Simms, V., Muldoon, K., and Towse, J. (2013). Plane thinking: Mental representations in number line estimation as a function of orientation, scale, and counting proficiency. *Journal of Experimental Child Psychology*, *115,* 468-480.

Sowinski, C., LeFevre, J., Skwarchuk, S., Kamawar, D., Bisanz, J, and Smith-Chant, B. (2015). Refining the quantitative pathway of the Pathways to Mathematics model. *Journal of Experimental Child Psychology, 131,* 73-93.

St Clair-Thompson, H., and Gathercole, S. E. (2006). Executive functions and achievement in school: Shifting, updating, inhibition and working memory. *The Quarterly Journal of Experimental Psychology, 59,* 745-759.

Swanson, H. L., and Sachse-Lee, C. (2001). Mathematical Problem Solving and Working Memory in Children with Learning Disabilities: Both Executive and Phonological Processes Are Important. *Journal of Experimental Child Psychology, 79,* 294-321.

Takahashi, A., and Green, D. (1983). Numerical judgments with Kanji and Kana. *Neuropsychologia, 21,* 259-263.

Van de Ven, S. H. G., Kroesbergen, E. H., Boom, J., and Leseman, P. P. M. (2012). The development of executive functions and early mathematics: A dynamic relationship. *British Journal of Educational Psychology, 82,* 100-119.

Vanbinst, K., Ghesquière, P., and De Smedt, B. (2012). Numerical magnitude representations and individual differences in children's arithmetic strategy use. *Mind, Brain, and Education, 6,* 129-136.

Visu-Petra, L., Cheie, L., Benga, O., and Miclea, M. (2011). Cognitive control goes to school: The impact of executive functions on academic performance. *Procedia Social and Behavioural Sciences, 11,* 240-244.

von Aster, M. G., and Shalev, R. S. (2007). Number development and developmental dyscalculia. *Developmental Medicine & Child Neurology, 49,* 868-873.

Vukovic, R. K., and Lesaux, N. K. (2013). The language of mathematics: investigating the ways language counts for children's mathematical development. *Journal of Experimental Child Psychology, 115,* 227-244.

Waber, D. P., Gerber, E. B., Turcios, V. Y., Wagner, E. R., and Forbes, P. W. (2006). Executive functions and performance on high-stakes testing in children from urban schools. *Developmental Neuropsychology, 29,* 459-477.

Wagner, R., Torgesen, J. K., and Rashotte, C. (2011). *Test of Word Reading Efficiency – Second Edition.* Austin, Texas: Pro-ed.

Wang, L. C., Tasi, H. J., and Yang, H. M. (2012). Cognitive inhibition with and without dyslexia and dyscalculia. *Research in Developmental Disabilities, 33,* 1453-1461.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI).* San Antonio, Texas: Harcourt Assessment Inc.

Whyte, J. C., and Bull, R. (2008). Number games, Magnitude representation, and number skills in pre-schoolers. *Developmental Psychology, 44,* 588-596.

Wiese, H. (2003). Whalen, J., Gallistel, C. R., and Gelman, R. (1999). Non-Verbal Counting in Humans: The Psychophysics of Number Representation. *Psychological Science, 10,* 130-137.

Wilson, K. M., and Swanson, H. L. (2001). Are Mathematics Disabilities Due to a Domain-General or a Domain-Specific Working Memory Deficit? *Journal of Learning Disabilities, 34,* 237-248.

Wolf, E. J., Harrington, K. M., Clark, S. L., and Miller, M. W. (2013). Sample size requirements for structural equation models an evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 73*, 913-934.

Wynn, K. (1998). Psychological foundations of number: numerical competence in human infants. *Trends in Cognitive Sciences, 2,* 296-303.

Xu, F., and Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74,* B1-B11.

Zhang, X., Koponen, T., Räsänen, P., Aunola, K., Lerkkanen, M.-K., and Nurmi, J.-E. (2014). Linguistic and spatial skills predict early arithmetic development via counting sequence knowledge. *Child Development, 85,* 1091–1107.