Children’s Arithmetic Development: Contributions of Symbolic and Nonsymbolic Magnitude Comparison

Sarah Elizabeth Watson

Submitted for the degree of PhD

University of York

Department of Psychology

September 2013

Abstract

This thesis aimed to explore the predictors of children's arithmetic development with a specific focus on magnitude comparison. Children were assessed in whole class groups in order to recruit a sample large enough to use structural equation modeling (Chapters 2, 4 and 5), while also assessing a subsample of children individually with computerised measures (Chapter 6). This thesis also aimed to explore children's development on the magnitude comparison tasks within the same group of children (Chapters 3 and Chapter 6 Study 1).

Chapter 2 first assessed the underlying latent factors that different comparison tasks may have in common. It was found that symbolic and nonsymbolic comparison tasks loaded on the same factor (magnitude comparison), whilst letter comparison formed a separate factor. Furthermore, children's magnitude comparison ability was found to be a concurrent predictor of their arithmetic achievement but letter comparison was not. The longitudinal analyses in Chapters 4 and 5 show how magnitude comparison ability was not a predictor of children's untimed arithmetic ability, or fluency at completing subtraction and multiplication problems either one or two years later. However, it was a significant predictor of addition fluency one year later. In comparison, number identification ability was found to be a consistent predictor of arithmetic achievement both concurrently and longitudinally.

Chapter 6 investigated whether the inconsistent findings regarding the importance of magnitude comparison ability was due to the methodology used to assess it. Computerised magnitude comparison tasks more akin to those in previous studies were individually presented to a subgroup of children that also completed the group based measures. Neither symbolic nor nonsymbolic comparison ability was found to predict later arithmetic achievement, whereas number identification was a significant predictor.

Finally in Chapters 3 and 6, it was found that children improved significantly over time on all of the magnitude comparison tasks presented.

Table of Contents

Appendices

List of Figures

List of Tables

Acknowledgements

Firstly I would like to thank my supervisor, Silke Göbel, for her constant support, guidance and encouragement during my PhD. Thank you for your time and patience over the past four years. I would also like to say thank you to Charles Hulme for his time teaching me to use Mplus and for his guidance with the data analysis. I would also like to acknowledge Arne Lervåg for his assistance with this too. Thank you also to the members of my research committee.

I have thoroughly enjoyed being part of the CRL, it has been an amazing environment to conduct my research within and I have learnt a lot from so many people. I would like to thank everyone for their support and friendship. Thank you to Maggie Snowling and Charles Hulme for initially giving me the opportunity to be part of this research group and to Paula Clarke for encouraging me to undertake a PhD. A special thank you goes to Silvana Mengoni and Faye Smith as I'm not sure I could have survived the final writing up phase without you!

I would also like to take the opportunity to thank Gary Jones for initially giving me the belief and confidence during the early stages of my career with my BSc project and Research Assistant post.

A big thank you to all of the schools and children (and their parents) who participated in this research, it would not be here without them. I am also grateful to the BBSRC for funding this research.

Finally, I am indebted to my family who have supported and believed in me all the way, a few words here are never going to be enough. Thank you to my mum, dad, and sister, you have kept me going when I wanted to give up and also provided just the right amount of distraction to keep me sane!

Author’s Declaration

I declare that the work presented within this thesis is my own work and has not been previously submitted to this or any other University for a degree or any other qualification. This research was supported by a Biotechnology and Biological Sciences Research Council (BBSRC) studentship awarded to the candidate.

Selected aspects of the research have been presented elsewhere:

Göbel, S. M., Watson, S. E., Lervåg, A., & Hulme, C. (2014). Children’s arithmetic development: it is number knowledge, not the approximate number sense, that counts. *Psychological Science.* Advance online publication. doi: 10.1177/0956797613516471

This paper is based on part of the data collected by SW (the candidate) and presented in Chapter 4. The paper was written by SGo and the data analyses in this paper are different to those presented in Chapter 4, they include additional SEM modelling carried out by CH and AL.

Göbel, S. M., Watson, S. E., Lervåg, A., & Hulme, C. Number comparison versus number identification: longitudinal predictors of growth in arithmetic in primary school. *Paper presented at TeaP, Vienna, March 2013.*

This conference paper was based on the data presented in above paper with additional data taken from Chapter 4.

Watson, S. E., & Göbel, S. M. Just choose the larger numerosity? Longitudinal predictors of arithmetic. *Paper presented at the BPS Developmental Psychology Conference, Glasgow, September 2012.*

Watson, S. E., Hulme, C., & Göbel, S. M. Just choose the larger numerosity: Concurrent predictors of arithmetic skill in 5 and 6 year olds. *Poster presented at the BPS Developmental Psychology Conference, London, September 2010.*

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

The research in Chapter 6 was aided by Emily Barrett and Laura Watts-Chapman who collected the nonsymbolic comparison data at Time 3 as part of their BSc Psychology dissertation project.

The dot counting and nonsymbolic comparison tasks used in Chapter 6 were initially developed by Kristina Moll in a project lead by Karin Landerl.

Chapter 1.

The Importance of Numerical Processing for Arithmetic Development

1.1. Overview

Competence in numeracy is important for everyday life. Whilst in areas such as reading there are clear early predictors of later reading success, we know less about longitudinal predictors of mathematics and arithmetic. There are four main skills that have been suggested to be predictors of arithmetic achievement in children: numerical processing/magnitude comparison (e.g. Durand, Hulme, Larking & Snowling, 2005), counting (e.g. Passolunghi, Vercelloni & Schadee, 2007), working memory (e.g. Bull, Epsy & Wiebe, 2008) and phonological awareness (e.g. Hecht, Torgessen, Wagner & Rashotte, 2001). There is also disagreement over the importance of some of these factors, in particular magnitude comparison. The main aim of this thesis is to investigate whether magnitude comparison tasks are reliable independent longitudinal predictors of growth in arithmetic. This chapter will begin with a summary of models of number representation, followed by a review of the literature of magnitude comparison tasks and in particular will then focus on literature about how performance on these measures relates to children's arithmetic skill.

1.2. Models of Numerical Processing

Several models have been proposed to explain numerical processing in adults (e.g. Numerosity code model, Zorzi & Butterworth, 1999; Zorzi, Stoianov & Umilta, 2005; Single semantic route model, McCloskey, 1992; Multi-route model, Cipolotti & Butterworth, 1995). The most commonly cited and empirically validated model that has also been most influential in the developmental literature is Dehaene's triple code model (Dehaene, 1992; Dehaene & Cohen, 1995).

1.2.1. Dehaene's Triple Code Model

The model proposes that numbers are represented and processed by three distinct codes: the *analogue magnitude code*, the *visual Arabic number form* and the *auditory verbal word frame* (see Figure 1.1 for a representation) (e.g. Dehaene, 1992). The codes are interlinked by paths that translate representations from one code to another.

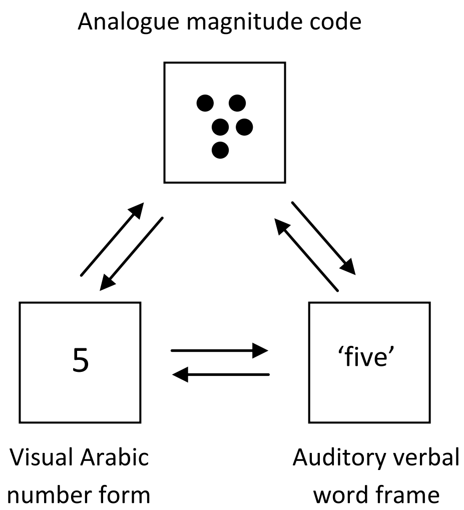


Figure 1.1. Representation of the triple code model.

The *analogue magnitude code* is a system that represents quantity or magnitude in an approximate way and is the only code that represents the meaning of a number (semantic representation of number) (Dehaene, 1992). This ability to represent magnitudes approximately is believed to be shared by animals and preverbal infants (e.g. Dehaene, 1992; 1997; Whalen, Gallistel & Gelman, 1999; Wynn, 1998) and this intuitive sense of number is often referred to as *the number sense* (Dehaene, 1997). Dehaene and his colleagues propose that magnitudes are represented by distributions of activation on an internal number line which is oriented in a left to right direction in Western participants (see Figure 1.2) (e.g. Dehaene, 1992; Dehaene, Dupoux & Mehler, 1990). Support for this spatial orientation has been found from parity (odd or even) judgement tasks where smaller numbers are responded to faster with the left than the right hand, and larger numbers are responded to more quickly with the right hand (Dehaene, Dupoux & Mehler, 1990; Dehaene, Bossini & Giraux, 1993). This pattern of results has also been found in children from 9 years old (Berch, Foley, Hill & Ryan, 1999). The representation of magnitudes follows a logarithmic scale and the signature of this model is its compression; magnitude representations are not evenly spaced along the number line, rather, larger numbers are represented more closely together than smaller numbers but with fixed variability (Dehaene, 1992; 1997).

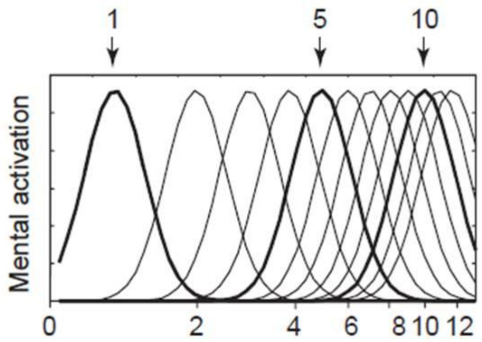


Figure 1.2. Logarithmic representation with fixed variability. Adapted from Feigenson, Dehaene and Spelke, 2004, p. 309.

It should be noted that Gallistel and Gelman (1992) propose an alternative conceptualisation of magnitude representation; they suggested that the mapping from numerical value to the magnitude representation is linear rather than logarithmic (see Figure 1.3 for a pictorial representation). In their model magnitude representations are noisy; the amount of noise is proportional to the size of the magnitude, therefore larger magnitudes have more noise (scalar variability) (Gallistel & Gelman, 1992; 2000). This pattern can be likened to a normal distribution, where each number has a mean magnitude with variability distributed around this mean like a standard deviation. Therefore larger means will also have a larger standard deviation so greater distribution and more variability in the magnitude representation. Whereas there is debate in how the magnitudes that represent numbers are positioned on this internal number line, both proposals result in the same pattern of effects being observed (discussed in detail below) and it has been suggested that more complex arithmetical skill builds upon this system (Dehaene, 1997; Gallistel & Gelman, 1992; Wynn, 1992).

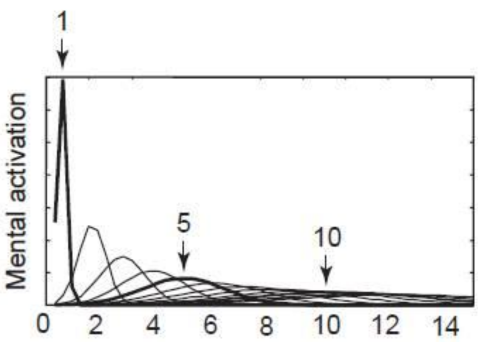


Figure 1.3. Linear representation with scalar variability. Adapted from Feigenson, Dehaene and Spelke, 2004, p. 309.

Humans and animals are thought to share an underlying system for representing number (the analogue magnitude code), whereas the following two codes are unique to humans. Dehaene (1992) termed the code that represents numbers verbally (e.g. twenty eight), the a*uditory verbal word frame*. This code draws on the general language processing system and carries out such tasks as verbal counting and multiplication tables. In the *visual Arabic number form* Arabic numerals are represented and manipulated as strings of digits (e.g. 28). This code is dedicated to numerical material and deals with multi-digit calculations and parity judgements.

Input procedures turn external stimuli into the notation specific representations, where numbers can then be transcoded into any of the other codes for the numerical procedure to take place. Similarly each representation has a specific output code (Dehaene & Cohen, 1995). With regards to solving arithmetic problems Dehaene and Cohen (1995; 1997) suggested that there are two possible routes (for other views see McCloskey, 1992; Cipolotti & Butterworth, 1995). A direct asemantic route which would be used when answering over learned calculations, particularly multiplication facts. This would involve the auditory verbal word frame as facts are stored as verbal associations in memory. The second is an indirect semantic route which would be used when solving subtraction and more complex addition problems. The operands are encoded as quantity representations which can then be manipulated in order to complete the problem.

1.2.2. Developmental Models of Numerical Processing

The adult models of numerical processing have inspired research into the development of these skills. Most developmental models are based on adult models and are currently more like descriptive frameworks that are in need of empirical validation.

Von Aster and Shalev (2007) proposed a four step developmental model of number acquisition (see Figure 1.4). The model is hierarchically organised with children moving through the stages with schooling and increasing working memory capacity. Similar to Dehaene's proposal they suggest that children have a preverbal core system of magnitude, Step 1, which represents cardinal magnitude and provides the meaning of number. In Step 2, the verbal number system, children learn the number words which are associated with numbers of objects. Step 3, the Arabic number system, occurs when children are at school and learn the symbolic representation of the magnitudes. Step 4, the mental number line, develops during the school years; this is ordinal and marked with Arabic notations. They propose that while the representation of cardinal magnitude is innate, the ordinal number line is acquired, and the early preverbal system that infants possess is different to the mental number line used for more complex mathematical reasoning in older children and adults.

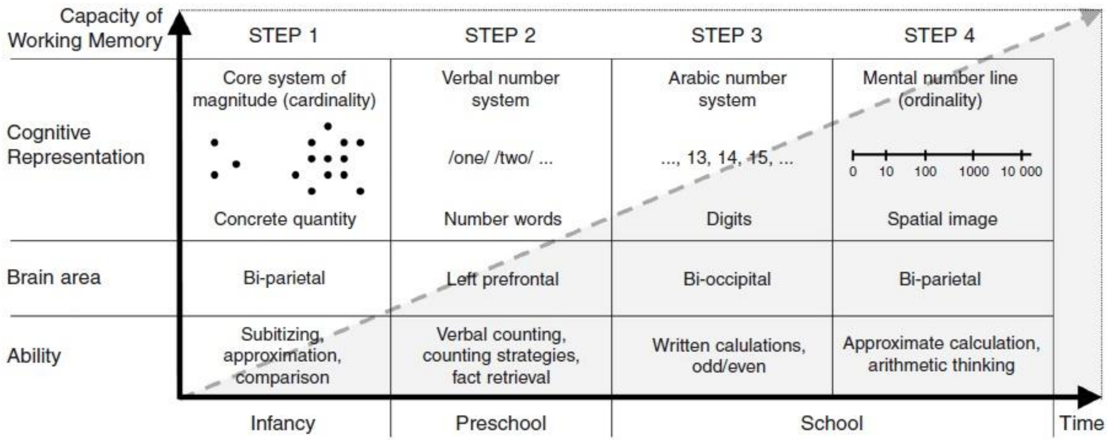


Figure 1.4. Von Aster and Shalev's four set developmental model of numerical cognition. From Von Aster and Shalev, 2007, p. 870.
Shaded area below broken line represents increasing working memory.

In contrast, Feigenson, Dehaene and Spelke (2004) proposed that infants, children and adults possess two core systems for representing number. The first system processes large numerosities and represents numerical magnitude approximately as a mental number line, similar to Dehaene (1992) and Gallistel and Gelman (1992). This system later becomes integrated with a symbolic number system. In addition to the large number system, a second system keeps track of small numbers of individual items and represents this information precisely. It is suggested to have a limit of about three in infants and could be the reason that in adults small numbers of items (1 to 4) are processed so quickly and accurately (although this remains controversial). In this thesis any nonsymbolic numerosities presented to children will be outside of the small range (i.e. will be larger than 4). The large approximate number system will therefore be the focus.

A common theme with all of the proposed models (both adult and developmental) is the assumption that humans possess a way to represent number approximately (nonsymbolic number processing), and that this is present even in the early stages of life. In addition, symbolic number processing using number words or Arabic digits has to be acquired. The following sections will present an overview of the empirical findings on which the models have been built.

The number line models presented suggest that when two numerosities are presented it would be more difficult to discriminate between them if they had a small numerical distance between them than it would if the numerical distance was large. This idea stems from the highly replicated effect first discovered by Moyer and Landauer (1967) - the *distance effect*. They found that when adults were asked to indicate which of two presented Arabic numerals was numerically larger, the time taken to make this decision depended on the numerical distance between the numbers. For example, when the distance between two numbers is large (e.g. 1 vs. 7) decision times are faster than when the distance between them is small (e.g. 4 vs. 6). This profile of results, with longer decision times for numerically close stimuli, is unlikely to be due to participants using a counting process as the opposite pattern of results would be expected, i.e. shorter reaction times for numerically close digits (counting up from the smaller). This distance effect has also been reflected in comparison accuracy; with accuracy performance increasing as the distance increases. Moyer and Landauer (1967) proposed that the decision process reflects the digits first being converted into analogue magnitudes, where the comparison is then made.

According to some models (Dehaene, 1992; Gallistel & Gelman, 1992), numerosities that have a smaller numerical distance between them would be represented closer together on the number line than would numerosities with a larger numerical distance between them. The variability of the representations of magnitude would therefore have greater overlap when they are closer together making it more difficult to discriminate between them and choose the larger.

Instead of focusing on the absolute numerical distance some recent studies have investigated the ratio difference between two numerosities and its effect on comparison performance (e.g. Barth, La Mont, Lipton, Dehaene, Kanwisher & Spelke, 2006). This takes into account the size of the numerosities to be compared as well as the numerical distance between them. Comparison accuracy has been found to decrease as the ratio between the numerosities decreases (bigger number/smaller number), in adults and children (e.g. Barth et al., 2006; Halberda, Mazzocco & Feigenson, 2008; Inglis, Attridge, Batchelor & Gilmore, 2011). With regards to the number line models it would be more difficult to discriminate between larger numbers than smaller numbers due to the proposals that larger numbers are represented more closely on the number line (Dehaene, 1992), or that larger numbers are represented with greater variability (Gallistel & Gelman, 1992).

1.3. Infants Preverbal Understanding of Number

To investigate whether infants were able to represent numerosities approximately, Xu and Spelke (2000) tested whether six month old infants could discriminate between two nonsymbolic numerosities (using 8, 12, and 16). They used a habituation technique which involved presenting infants with displays depicting the same numerosity (e.g. 8 dots), when looking time declined infants were then presented with test displays alternating between the same numerosity (e.g. 8) and a new numerosity (e.g. 16). It was found that infants looked longer at the display showing the new numerosity than at the display they had been habituated to, but this was only when the ratio between them was 1:2 (8 vs. 16) and not when it was reduced to 2:3 (8 vs. 12). The results suggest that infants possess approximate representations of number as the findings cannot be accounted for by mechanisms of object-based attention (Xu & Spelke, 2000). It also appears that infants were only able to discriminate between two numerosities when the ratio between them was large (1:2).

This finding has been supported by research where numerosities were presented as auditory sequences; this also provides support for the proposal that an abstract representation of number is present in infants (Lipton & Spelke, 2003). Lipton and Spelke (2003) investigated the ratio at which 9 month old infants could discriminate between presented sounds and found that they were able to discriminate when the ratio was 2:3 but not when the ratio was 4:5. Taking these two studies together it shows that infants' ability to discriminate between numerosities increases with age.

1.4. Magnitude Comparison

To investigate how number may be represented in older children and adults the more active task of magnitude comparison is often used. This technique involves presenting two nonsymbolic numerosities (usually as dots or squares) or Arabic numerals and discriminating between them by choosing the array/number that is the largest in magnitude (i.e. the most dots/bigger number). The distance or ratio between the items is varied systematically to induce a distance or ratio effect.

A point to note when reviewing the literature on the effect of distance/ratio on magnitude comparison accuracy and reaction times is that the method of data analysis often varies. When the numerical distance is varied (often 1 to 6) this is often entered as a factor in an ANOVA when analysing both accuracy (as percentage correct) and reaction times (e.g. Holloway & Ansari, 2008). An alternative method is to quantify individual differences in the effect of distance; most commonly on reaction times as accuracy performance is generally very high on these tasks. This has been quantified in two different ways: calculating a distance effect, this usually involves subtracting the mean reaction time for comparisons with a large distance (e.g. 5 and 6) from the mean reaction times for comparisons with a small distance (e.g. 1 and 2) and then dividing this figure by the reaction times for comparisons with a large distance (5 and 6) $\left(\frac{RT_{small} - RT_{large}}{RT_{large}}\right)$ (see Holloway & Ansari, 2009; Mundy & Gilmore, 2009). The second method is to calculate a regression slope with change in reaction times predicted by distance (see De Smedt, Verschaffel & Ghesquière, 2009). In relation to the number line models a larger distance effect and a steeper (or larger) slope would therefore suggest that an individual had greater overlap in their internal representations of numerical magnitudes. An increasingly popular technique used when investigating performance accuracy on nonsymbolic comparison tasks is to calculate a Weber fraction (see Halberda et al., 2008; Piazza et al., 2010; Piazza et al., 2004). This idea is based on psychophysical models (more detail can be found in Chapter 6) and it is proposed that the size of the Weber fraction represents the acuity of magnitude representations; the smaller the Weber fraction the greater acuity of the representations within the approximate number system.

1.4.1. The Nonsymbolic Ratio Effect

As would be expected, it has been found that children's overall accuracy on these nonsymbolic comparison tasks improves with age (e.g. Bonny & Lourenco, 2013; Halberda & Feigenson, 2008; Mussolin, Nys, Leybaert & Content, 2012; Soltesz, Szűcs & Szűcs, 2013). However, it should be noted that these studies are cross-sectional rather than longitudinal.

Halberda and Feigenson (2008) found that the development in discrimination ability continued through childhood. They tested children aged 3, 4, 5 and 6 years old, and compared their performance with adults. They were presented with two nonsymbolic numerosities (they used pictures of objects, e.g. balls and trains, ranging from 1 to 14 items) and varied the ratio between them (1:2, 2:3, 3:4, 4:5, 5:6, 6:7, 7:8, 8:9, 9:10). A significant effect of age was found with accuracy performance at distinguishing between the two numerosities increasing with age. It was found that the ratio at which each age group was successfully able to distinguish between the numerosities improved with age, from 2:3 for the 3 year old children to 6:7 for the 6 year old children, whereas adults were able to compare ratios of 9:10. These whole number ratios were translated from a Weber fraction that was calculated using a psychophysical model (based on Pica et al., 2004), which is an alternative way to conceptualise accuracy performance (more detail can be found in Chapter 6).

Piazza et al. (2010) also carried out a study designed to investigate the development of number sense. To gain an estimate of the precision of each individual's internal number representations they calculated a Weber fraction. Participants were always presented with the numerosity 16 (or 32) and a second array that was smaller or larger, for example when the numerosity 16 was presented it was paired with either 12, 13, 14, 15, 17, 18, 19, or 20 (paired with 32 were double these values). Their findings were consistent with Halberda and Feigenson (2008) in that the size of the mean Weber fraction reduced significantly with age group; kindergarteners (mean age = 5;03) = 0.34, 8 to 12 year olds = 0.25, and adults = 0.15.

Recently there has been a wealth of studies investigating the development of the Weber fraction in children (aged from 3 years) right up to adulthood, showing that the size of the Weber fraction decreases with increasing age (e.g. Halberda et al., 2012; Mussolin, Nys, Leybaert & Content, 2012; Libertus, Feigenson & Halberda, 2013; Sasanguie, Göbel, Moll, Smets & Reynvoet, 2013). This suggests that the process of refinement of the approximate number system continues through development. However, the majority of these studies are cross-sectional and therefore longitudinal studies exploring this development within the same children are needed.

1.4.2. The Nonsymbolic Distance Effect

In line with the findings of Moyer and Landauer (1967), an effect of distance has also been observed when comparing nonsymbolic stimuli. When people are asked to indicate which of two numerosities is numerically larger accuracy increases as the distance between numerosities increases, while reaction times decrease (e.g. Holloway & Ansari, 2009). This pattern has also been observed in children of different age groups: 6 to 7 years old (Mundy & Gilmore, 2009), 6 to 8 years old (and also adults, Holloway & Ansari, 2008), and 10 to 11 years old (Mussolin, Mejias & Noël, 2010).

Age related changes have also been found on response times. Holloway and Ansari (2008) investigated the effect of distance on reaction times of 6, 7 and 8 year old children and adults. Participants were presented with two nonsymbolic numerosities ranging from 1 to 9, with the numerical distances between them ranging from 1 to 6. Due to possible age related changes in speed of processing (Kail, 1991; Kiselev, Espy & Sheffield, 2009), basic reaction time was used as a covariate in the analysis. Reaction times were found to reduce significantly with age, and more importantly a decrease in the distance effect was found with increasing age.

Sasanguie, De Smedt, Defever & Reynvoet (2012) also investigated the effect of distance on children's comparison ability. Four age groups of children were chosen; kindergarten (mean age = 5.6 years), first grade (mean age = 6.7 years), second grade (mean age = 7.6 years), and sixth grade (mean age = 11.6 years). As in Holloway and Ansari (2008), the numerosities ranged from 1 to 9 but the numerical distance between them ranged from 1 to 5. Children's reaction times were found to decrease with development (although they did not control for speed of processing). They also used an alternative way to quantify individual differences in the effect of distance on children's comparison performance; it was represented by the gradient of a regression slope. A linear regression was calculated on each individual's reaction times, in which reaction time on the task was predicted by distance (a method used by Fias, Brysbaert, Geypens & d'Ydewalle, 1996), speed of processing differences are therefore taken into consideration. A steeper slope would therefore be comparable to a larger distance effect. They found that the gradient of this slope decreased with increasing grade which suggests that the effect of distance decreased with increasing age.

1.4.3. The Symbolic Distance Effect

As reported earlier, Moyer and Landauer (1967) first identified the distance effect in adult's comparison of Arabic digits. This finding that as the distance between two numbers increases the time taken to respond decreases has also been replicated in children (e.g. De Smedt et al., 2009; Sekuler & Mierkiewicz, 1977). Accuracy performance has also been found to increase with increasing distance (e.g. De Smedt et al., 2009).

The age related changes in nonsymbolic comparison performance are also reflected when symbolic stimuli are compared. When asked to indicate which of two digits is numerically larger, the time taken to make this decision decreases with age (e.g. Sasanguie, De Smedt et al., 2012; Sekuler & Mierkiewicz, 1977). Sekuler and Mierkiewicz (1977) found that mean reaction times for children in kindergarten and first grade were significantly different to each other and also slower than children in both fourth and seventh grade and adults. There was no significant difference in reaction times between the three older groups. Sekuler and Mierkiewicz (1977) proposed that numerical representations change with development and that younger children may have more compressed representations and/or more variability in their representations. The longer mean reaction times for the two younger groups of children in Sekuler and Mierkiewicz's (1977) study could be due to age related speed of processing differences (Kail, 1991; Kiselev, Espy & Sheffield, 2009). Holloway and Ansari (2008) addressed this issue by controlling for basic reaction time in their analysis and still found that reaction times decreased with increasing age.

Not only have changes in overall response times been found, the distance effect also has also been found to decrease with age (e.g. Holloway & Ansari, 2008; Sasanguie, De Smedt et al., 2012; Sekuler & Mierkiewicz, 1977). As with the nonsymbolic comparison task used in their study, Sasanguie, De Smedt et al. (2012) found that the gradient of the regression slope (change in RT predicted by distance) decreased with increasing grade. Meaning that as children get older they are less affected by the distance between the two symbolic digits they are comparing.

To my knowledge only one longitudinal study exists that investigates the differences in symbolic comparison ability as a function of age. Reeve, Reynolds, Humberstone and Butterworth (2012) carried out a longitudinal study testing children at 6, 7, 8.5, 9, 9.5, 10 and 11 years old (mean ages at each testing point). As with the previous symbolic comparison tasks they varied the distance between the numbers (all combinations of all numbers 1 to 9). In line with the cross-sectional studies a decrease in reaction times was found with increasing age. The time taken to compare each ratio also decreased at successive ages. Furthermore, a regression slope was calculated on the data and they found that the effect of distance on reaction times decreased with age.

These findings show that through development the ability to compare two numerical stimuli increases, whether this is symbolic Arabic numerals or arrays of nonsymbolic items. This pattern is found for accuracy performance, for example the decrease in the size of the Weber fraction, and also in the speed with which comparisons are made, with both overall reaction times and the effect of distance decreasing with age. These patterns of results are suggested to reflect the refinement in a child's internal representation of number (e.g. Halberda & Feigenson, 2008; Sekuler and Mierkiewicz, 1977). However, it is important to note that the majority of data so far is cross-sectional, what is needed is longitudinal data for the same children.

1.4.4. The Relationship between Symbolic and Nonsymbolic Comparison Ability

Whilst some researchers investigate the representation of number using either symbolic or nonsymbolic comparison tasks it may be important to assess both within the same group to gain further understanding about the nature of this system. If performance on both nonsymbolic and symbolic comparison tasks (perhaps more specifically the distance effect and the Weber fraction) are used as an estimate of the acuity of the internal representation of numerical magnitude then it would be expected that performance on the two types of task is related. If not, then this raises the question about what performance on these tasks is actually measuring. There is little previous research that investigates this question specifically but some studies have used both a symbolic and nonsymbolic comparison task to index magnitude comparison ability. The following section will review both of these in turn.

Maloney, Risko, Preston, Ansari and Fugelsang (2010) carried out a study with the specific aim of investigating the validity of the distance effect by comparing participants' distance effects on multiple variants of magnitude comparison tasks. Adults carried out two types of symbolic comparison task (comparing two Arabic digits in the range of 1 to 9 but excluding 5, and comparing one Arabic digit to a standard which was 5) and a nonsymbolic comparison task (comparing two numerosities between 1 and 9 but excluding 5). A distance effect was calculated on both reaction times and accuracy performance but this differs to the method reported previously (performance on distance 1 – distance 4). Only the cross comparison results will be reported here. No associations were found between the symbolic and nonsymbolic distance effects calculated using reaction times or accuracy scores in experiment one. In a second experiment a significant relationship was found between the distance effects calculated on accuracy scores for the symbolic lower/higher than 5 task and the nonsymbolic task, although it was weak in strength (r = .31). Due to the lack of relationships between the magnitude comparison tasks they concluded that the underlying mechanisms that give rise to the symbolic and nonsymbolic distance effects are different.

To explore the relationships between measures that have been suggested to assess the acuity of the approximate number system, Gilmore, Attridge and Inglis (2011) also administered multiple symbolic and nonsymbolic tasks to a group of adult participants and calculated multiple measures of performance (accuracy, Weber fraction and distance effect). They included magnitude comparison tasks using small numerosities (1 to 9) as well as large numerosities (nonsymbolic items ranged from 9 to 70; symbolic items ranged from 10 to 69). No relationship was found when comparing performance on the two large numerosity tasks (accuracy or Weber fraction), or when comparing performance on the two small numerosity tasks when the effect of distance on reaction times was compared. However, significant relationships were found between accuracy performance on the small numerosity comparison task but again these were quite weak in strength (mean accuracy: $r = .31$, distance effect: $r = .31$).

These two studies involved adult participants, to the best of my knowledge there is currently no research that has specifically investigated the relationship between performance on a symbolic and nonsymbolic task (i.e. validity) in children. However, there are some studies that have included both symbolic and nonsymbolic comparison tasks in their design and have reported the relationship between them (there are also many studies where this relationship is not reported).

First, concentrating on the effect of distance on comparison performance, Holloway and Ansari (2009) investigated the distance effect in a group of children aged 6 to 8 years old. They presented 1 to 9 items and calculated the distance effect in the more common way (described earlier). They found no relationship between the symbolic and nonsymbolic distance effects when it was calculated using accuracy scores or on reaction times (RTs). This finding was replicated by Lonnemann et al. (2011) in a group of children ranging from 8 to 10 years old; they calculated the distance effect on RTs using a smaller range of items (1 to 6) and distances (1 to 4). Sasanguie, De Smedt et al. (2012) also presented stimuli ranging from 1 to 9 but only with a maximum difference of 5. They calculated a regression slope to represent the effect of distance on RTs (adjusted to represent both speed and accuracy) but still found that there was no relationship between the two tasks in children ranging from 5 to 11 years old (partial correlations controlling for grade). These findings were again echoed in a study including both typically developing children and children with mathematical difficulties; the distance effects calculated using reaction times were not related (Mussolin, Mejias & Noël, 2010).

With regards to overall performance rather than the effect of distance, the findings look somewhat different. Holloway and Ansari (2009) found a significant strong association ($r = .73$) between mean reaction times on the two tasks, whereas there was no relationship between accuracy performance on the two tasks. Mussolin et al. (2010) also found a significant correlation between mean RTs on a symbolic and nonsymbolic comparison task (*r* = .74), but the correlations investigating accuracy performance were not reported. Sasanguie, De Smedt et al. (2012) used RTs adjusted to reflect both speed and accuracy performance and again found a significant relationship between performance on the two tasks (*r* = .66).

In summary, where relationships between the two types of magnitude comparison task are found they are often weak in strength, except for the relationships between mean reaction times. It is possible that the significant correlations observed between reaction times reflect speed of processing skill rather than comparison ability. The finding that the distance effects derived from symbolic and nonsymbolic tasks are not related raises the question of whether these two types of magnitude comparison task are in fact tapping the same underlying construct. Alternatively, the method used to calculate a distance effect (typically a difference score or the difference sore divided by the RT for large distance comparisons) may not be the best method for categorising the effect. Differences scores can be unreliable (Peter, Churchill & Brown, 1993) and children's speed of processing will also influence them. Although Sasanguie, De Smedt et al. (2012) also found that there was no relationship between symbolic and nonsymbolic distance effects when the more advantageous method of conducting a regression slope was used. It should also be noted that in a large number of the studies presented above (Holloway & Ansari, 2009; Maloney et al., 2010; Mussolin et al., 2010; Sasanguie, De Smedt et al., 2012) children were comparing small numbers of items. The action of enumerating numerosities in the range of 1 to 3 has been found to be different to that of larger range numerosities (4 and beyond), both in adults (Mandler & Shebo, 1982), and children (Landerl, Bevan & Butterworth, 2004). The fast and accurate report of small items has been termed 'subitizing' (Kaufman, Lord, Reese & Volkmann, 1949). While children are explicitly told not to count when carrying out comparison tasks, if the subitizing process is automatic, then this may still contribute to the lack of relationship. It is therefore important that future research includes both symbolic and nonsymbolic comparison measures and examines the relationship between them when investigating the representation of number. It would also be advantageous to use nonsymbolic numerosities outside of this small range.

1.5. The Relationship between Magnitude Comparison Ability and Arithmetic Skill

Before the relationships between magnitude comparison ability and arithmetic achievement are reviewed it is important to note that variability in performance on tasks is often not acknowledged and data is treated as a group. When investigating the relationship between tasks individual differences in performance are needed and expected but often rarely discussed. There can be large individual differences in children's arithmetic achievement (see Dowker, 2005) and many studies have identified children with different arithmetic and mathematical skill levels (e.g. Geary, Hoard, Byrd-Craven, Nugent & Numtee, 2007; Landerl, Bevan & Butterworth, 2004). Large variability in performance on magnitude comparison tasks has also been observed with regards to the distance effect (e.g. Reeve et al., 2012) and the size of the Weber fraction (e.g. Halberda et al., 2008; Inglis et al., 2011).

The magnitude comparison tasks, and the effects observed, are suggested to assess the proposed internal representation of numerical magnitude. It has been proposed that this approximate number system may be a precursor for more complex arithmetic skill to build upon (Ansari & Karmiloff-Smith, 2002; Barth, La Mont, Lipton & Spelke, 2005). If this indeed is the case then a relationship would be expected with formal arithmetic achievement. The exploration of the relationship between individual differences in children's magnitude comparison ability and arithmetic/mathematics performance has recently received greater attention. It is acknowledged that there are domain general predictors of children's arithmetic achievement, for example, working memory (e.g. Bull, Espy & Wiebe, 2008), and there is clear evidence that counting ability is a domain specific predictor (e.g. Passolunghi, Vercelloni & Schadee, 2007), taking this into account the focus will remain on magnitude comparison ability and the approximate number system as a possible predictor of children's arithmetic achievement. The concurrent literature investigating the relationship between symbolic comparison and arithmetic will be reviewed first, followed by the relationship with nonsymbolic comparison. As these studies cannot tell us anything about the direction of the relationship, longitudinal studies investigating the relationship will then be reviewed.

First, it is important that the distinction between arithmetic and mathematics is made clear. Mathematics is the study of (relationships among) number, quantity, shape and space, whilst arithmetic is a branch of mathematics, and is concerned with numerical calculations. Therefore the term arithmetic will be used when the task assesses basic number skills (e.g. identifying and writing numbers and counting) and basic calculations involving addition, subtraction, multiplication and division and the term mathematics will be used when a task also includes problems outside of basic calculations, for example geometry, algebra.

1.5.1. Concurrent Relationships between Magnitude Comparison and Arithmetic Skill

1.5.1.1. Symbolic comparison and its relationship with arithmetic skill.

An early study which examined the concurrent relationship between symbolic comparison performance and arithmetic achievement was a large scale study by Durand et al. (2005). They included a digit comparison task alongside other possible predictors, when investigating the cognitive foundations of arithmetic and reading skills in children aged 7 to 10 years old. They found that digit comparison ability (Beta = .35), alongside verbal ability (Beta = .44), was a significant unique predictor of arithmetic performance (explaining 42% of variance in total). Digit comparison was not a predictor of children's reading skill, suggesting that this task is tapping something specific to arithmetic skill and not just associated with general attainment. The arithmetic outcome measure used in this study included both speeded (addition and subtraction) and untimed calculation tasks. The digit comparison task was a simple group presented paper and pencil task which required children to complete as many comparisons as they could within 30 seconds. The digits being compared in the task only varied by a distance of one or two so does not allow for the investigation of the effect of distance on children's comparison ability. Therefore, it can only be concluded from this study that some aspect of children's symbolic comparison is predictive of their arithmetic achievement. However, this study was the first to examine and find a relationship between digit comparison ability and arithmetic achievement, whilst controlling for children's general cognitive ability in the analysis. The method used in the study was novel (i.e. large group testing) and the metric obtained, number of items compared within 30 seconds, differs from that usually seen in the literature, therefore further research was needed to investigate this relationship.

As it has been found that there is an effect of distance on children's symbolic comparison response times and accuracy, it is proposed that individual differences in this distance effect could be related to children's arithmetic ability. Holloway and Ansari (2009) investigated this relationship in a group of 6, 7 and 8 year olds. They varied the distance between digits systematically to investigate its effect on both accuracy and response time. They quantified individual differences in the size of the distance effect using children's reaction times and found that it explained unique variance (4.9%) in arithmetic fluency scores (mathematics fluency subtest), over and above age, mean RT, the nonsymbolic distance effect (see section 1.5.1.2 below) and word reading ability. Children with larger distance effects gained lower fluency scores. There was also a weak but significant association with children's calculation scores (*r* = -.22), however the distance effect did not contribute unique variance to those scores. The two arithmetic measures were subtests from the Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew & Mather, 2001), mathematics fluency requires children to answer as many single digit addition, subtraction and multiplication problems as they can in three minutes, whereas the calculation subtest contains calculation problems that increase in difficulty with no time limit. As arithmetic fluency and symbolic comparison both require speeded access of magnitude information from Arabic numerals, the stronger relationship with fluency may reflect this process. No relationship was found between any of the arithmetic measures and either overall accuracy or the effect of distance on children's accuracy performance.

Mundy and Gilmore (2009) also found that individual differences in the size of the distance effect (on reaction times) were significantly related to performance on a test of school mathematics in children who were on average 7 years old (*r* = -.52), replicating the findings of Holloway and Ansari (2009). The test used in this study could be divided into two sections the first assessed knowledge of symbolic numbers (for example, identifying the largest number in a set), while the second involved calculation problems. The strength of the relationship was similar when only the calculation section of the test was used (*r* = -.54). In contrast to Holloway and Ansari (2009), significant relationships were also found between children's arithmetic and accuracy on the comparison task; children who were overall more accurate on the task were also found to have higher arithmetic scores (*r* = .52; calculation scores only *r* = .53). Accuracy at comparing digits was found to be a significant predictor when entered first into the regression analyses and when entered after accuracy performance on a nonsymbolic version of the task (15.5% of variance explained). Unlike Holloway and Ansari (2009) the authors did not present regression analyses to explore the predictive relationship of the distance effect (on reaction times) even though there was a significant association with arithmetic. It is worth noting that the correlations between the comparison task and arithmetic achievement were stronger in this study than in Holloway and Ansari (2009). As the comparison tasks were the same in both studies, and the children taking part were a similar age, this raises a question about the measure used to assess children's arithmetic ability and whether this could result in differences in across studies.

Budgen and Ansari (2011) found similar results to Holloway and Ansari (2009) using the same comparison task and arithmetic measures (WJ-III) in a group of first and second grade children (mean age = 7 years 4 months). Both the ratio (rather than distance) effect and also the regression slope were calculated as measures of the effect of ratio on children's comparison performance. In this study a smaller distance between numbers would result in a larger ratio difference (e.g. 6 vs. 8 = 0.75, whereas 2 vs. 8 = 0.25). Significant associations were found (when controlling for age) with fluency scores (slope r = -.39; ratio effect r = -.25) and with calculation scores (slope r = -.27). However, no significant relationship was found between the ratio effect and calculation scores (r = - .15 ns), mirroring the weaker association and the lack of contribution to calculation scores found in Holloway and Ansari (2009). In a set of regression analyses controlling first for age and basic reaction times the slope was found to predict significant unique variance in both fluency (8%) and calculation scores (3%). Calculating a regression slope to represent the effect of distance or ratio might therefore be a better predictor of children's arithmetic ability than the distance effect.

The finding that the symbolic comparison regression slope is significantly related to arithmetic and mathematics achievement has been replicated. Sasanguie, De Smedt et al. (2012) also calculated a regression slope to represent the effect of distance on children's symbolic comparison performance and assessed mathematics with a curriculum-based standardised achievement test from the Flemish Student Monitoring System (Dudal, 2000). They included a large age range of children (kindergarteners, first, second, and sixth graders) and calculated partial correlations controlling for grade. Significant associations with mathematics were found with both adjusted reaction times (r = -.24) and the slope (r = .22). They investigated these relationships further and found that for reaction times the association with mathematics changed over grades, with only a significant relationship in kindergarteners, whereas for the slope, significant associations were found for all ages. Unfortunately no regression analyses were carried out so its predictive contribution is unknown. A stronger relationship was found with the same mathematics measure (r = .41) in a more concentrated age range of children (mean age = 8 years 10 months) (Vanbinst, Ghesquière & De Smedt, 2012). As the focus of their study was children's arithmetic strategy use, the exploration of the variance explained in children's general mathematics achievement was not reported (though the slope predicted 12% of variance in children's fact retrieval frequency).

In a group of slightly older children (7;06 to 11;06), Landerl and Kölle (2009) found very weak associations between the symbolic comparison distance effect (indexed as the slope) and arithmetic scores. Timed arithmetic measures were used which assessed the different numerical operations separately. Although all associations are weak the results suggest that there might be some variation across the operations (addition: r = .13, p < .05;

subtraction: $r = .11$; $p < .06$; multiplication: $r = .03$, *ns*). This was a large scale representative study and included some children with very poor arithmetic ability, 262 typically developing children and 51 children with dyscalculia (so it may even be expected that the correlations are over-estimates of the relationships).

1.5.1.2. Nonsymbolic comparison and its relationship with arithmetic skill.

Whereas the relationship between symbolic comparison and children's arithmetic ability has generally been supported, the relationship between nonsymbolic comparison and arithmetic is less clear. As noted previously, researchers generally adopt one of two methods when including nonsymbolic comparison tasks in their research. Either to present items in the range of 1 to 9 and manipulate the effect of distance to match a symbolic version that is also being administered (Holloway & Ansari, 2009; Mundy & Gilmore, 2009; Sasanguie, De Smedt et al., 2012; Vanbinst et al., 2012), or use larger numerosities and manipulate the ratio difference between them (Bonny & Lourenco, 2013; Fuhs & McNeil, 2013; Gilmore et al., 2013; Inglis et al., 2011; Libertus et al., 2011; 2013). When varying the ratio between numerosities, the analysis tends to focus on accuracy scores and often a Weber fraction is calculated (to represent the acuity of the approximate number system), whereas when smaller numerosities are used the analysis is performed on reaction time data (and sometimes explores accuracy data). The aim of this next section will be to give an overview of all of these measures.

Holloway and Ansari (2009) found that the size of the nonsymbolic distance effect did not explain variance in children's arithmetic scores (calculation or fluency subtests from the WJ-III) and this was the case whether using reaction time or accuracy measures (after controlling for age, symbolic comparison processing speed, symbolic comparison distance effect and reading). The only significant relationship between nonsymbolic comparison performance (calculated using overall accuracy, distance effect on accuracy scores, mean comparison RT, distance effect on RTs) and arithmetic, was between mean comparison RT and mathematics fluency scores, however this was weak in strength ($r = -.21$). They proposed that as a relationship with arithmetic achievement was found only with the symbolic comparison task, that this reflects the ability to access numerical magnitude information from symbols.

Mundy and Gilmore (2009) found no relationship between children's mathematics scores and the nonsymbolic distance effect on response times but there was a relationship with overall accuracy ($r = .35$, $p = .05$). Accuracy performance on the task did predict variance in children's mathematics scores when it was entered first into the regression model (12.2%) but not over and above performance on the symbolic version of the task. Therefore the relationship between nonsymbolic comparison performance and mathematics may be mediated by performance on symbolic comparison tasks.

Similar results were observed in two further studies that administered both symbolic and nonsymbolic comparison tasks. While relationships were found between mathematics ability and symbolic comparison performance, no corresponding relationships were found with nonsymbolic comparison performance (De Smedt & Gilmore, 2011; Sasanguie, De Smedt et al., 2012; Vanbinst et al., 2012).

The nonsymbolic tasks used in the above studies include numerosities ranging from 1 to 9. Often this is because a symbolic version is also presented and so that the two tasks are analogous. It was highlighted earlier that the action of enumerating small range numerosities (1 to 3) is different to that of larger range numerosities (4 and beyond) (Landerl et al., 2004; Mandler & Shebo, 1982). Individuals are able to enumerate small numbers of items fast and accurately (Kaufman et al., 1949), whilst when presented with larger arrays the response is based on counting (Mandler & Shebo, 1982). Therefore it is possible that the processes involved when comparing small numbers of items differs to that when comparing larger number of items (Revkin, Piazza, Izard, Cohen & Dehaene, 2008).

An alternative way to estimate the precision with which numerical magnitudes are represented is to calculate a Weber fraction. Libertus et al. (2011) investigated whether ANS acuity in young children who had received little formal instruction in mathematics would be related to their arithmetic achievement. Children were recruited from preschools and had an average age of 4 years 2 months (range = 2 to 6 years old). The comparison task involved comparing two arrays (one blue, one yellow) with the number of dots in each ranging from 4 to 15, and the ratio between the two being 1:2, 2:3, 3:4, or 6:7. Both children's accuracy and reaction times were recorded enabling three measures of performance: accuracy as percent correct, accuracy as the Weber fraction, and reaction time (this included times from both correct and incorrect comparisons). To assess children's arithmetic achievement the Test of Early Mathematics Ability (TEMA-3; Ginsburg & Baroody, 2003) was administered which includes tasks such as reading numbers, counting, and addition and subtraction calculation problems. Children who were more accurate and faster on the comparison task were found to have higher arithmetic scores (accuracy (% correct) $r = .42$; Weber fraction $r = -.26$; RT $r = -.28$), moreover, all three measures were found to predict variance in arithmetic with age and vocabulary knowledge controlled (RT and accuracy: RT = 5%, accuracy = 13% of variance explained; RT and Weber fraction: RT = 8%, Weber fraction = 6% of variance explained).

In the same sample of children re-tested on the same measures, these significant associations were found to remain six months later (Libertus et al., 2013). Again, children who were more accurate and faster on the comparison task were found to have higher arithmetic scores (accuracy (% correct) r = .52; Weber fraction r = - .42; RT r = -.36) and all three measures were found to be significant predictors (ranging from 4 to 18% of variance explained).

Even when using the same arithmetic outcome measure as previous studies (TEMA-3) and children of a similar age range (3 to 5 years old), the findings are not always consistent. Fuhs and McNeil (2013) found no relationship between accuracy performance (percentage correct) and arithmetic achievement. The nonsymbolic comparison task used did differ to those used in the majority of studies: it was not presented to children on a computer and a larger range of numerosities, with some within the subitizing range (1 to 30), were presented. This will be explored more in Chapter 4 but does highlight the fact that the relationship between performance on nonsymbolic comparison tasks and arithmetic may depend on the stimuli, presentation or the data analysis of the task.

To explore whether the relationship between number acuity and arithmetic was the same in children (mean age = 8 years 6 months) and adults, Inglis et al. (2011) calculated a Weber fraction for both groups and assessed arithmetic using the calculation subtest from the WJ-III. After controlling for nonverbal ability and age, it was found that there was a significant relationship between the size of the Weber fraction and arithmetic in children (r = -.55) but not in adults. The adults completed additional subtests (including mathematics fluency and more applied problems) from the WJ-III but there were no significant associations with any of these measures (all correlations were weak in strength). The comparison task completed by the adults included larger numerosities (9 to 70) than the task presented to the children (5 to 22), but with similar ratios, to make the task more difficult and to avoid ceiling effects. This therefore led to a wide range in the size of the Weber fraction in both groups. The authors speculate that the approximate number system does play a role in the early development of the understanding of number but that after mastering these early skills other factors, such as strategy choice or working memory capacity, may then lead to individual differences in arithmetic ability. What is therefore needed is a longitudinal study that explores both the concurrent and longitudinal predictive relationships between number acuity and arithmetic, whilst also assessing other possible factors that could lead to individual differences.

In contrast, Lyons and Beilock (2011) did find a significant relationship between individual differences in the size of the Weber fraction and (complex) mental arithmetic problems in a group of adults (r = -.34). However, this was mediated by individuals' symbolic number ordering ability (deciding whether or not triads of Arabic digits were in increasing order). This finding is similar to that of Mundy and Gilmore (2009) who found that children's accuracy on the nonsymbolic comparison task was related to their arithmetic achievement but did not predict variance in arithmetic scores once symbolic comparison ability was controlled.

1.5.1.3. Differences in the relationship with arithmetic between symbolic and nonsymbolic comparison ability.

Many of the papers presented include only one measure of magnitude comparison ability, either a symbolic or nonsymbolic comparison task. Recently more studies have included both (De Smedt & Gilmore, 2011; Holloway & Ansari, 2009; Landerl & Kölle, 2009; Mundy & Gilmore, 2009; Sasanguie, De Smedt et al., 2012; Vanbinst et al., 2012) which allows for the investigation of whether performance on either or both is related to arithmetic achievement. Although not all studies present all possible indexes to represent performance on the tasks (overall accuracy, mean reaction time, distance effect on accuracy scores, distance effect on reaction times, regression slope on RTs to represent the effect of distance), the most consistent finding of a relationship with arithmetic achievement is with symbolic rather than nonsymbolic comparison measures. Holloway and Ansari (2009) suggested that it is individual differences in retrieving or processing magnitude information from symbols that are predictive of arithmetic achievement.

1.5.1.4. Summary of the concurrent relationships between magnitude comparison and arithmetic ability.

In summary there are more consistent findings of a relationship between symbolic comparison performance and arithmetic achievement than with nonsymbolic comparison performance and arithmetic. Although when using symbolic comparison tasks the results are still not clear cut; where both the results of accuracy and speed are provided it is speed with which comparisons are made and the effect of distance on decision times which show the most promising results. This is possibly due to the fact that individuals make only small numbers of errors when comparing symbolic stimuli, which leads to reduced variance. Compared to symbolic comparison tasks, there are more differences between the nonsymbolic versions presented in the different studies, for example the range of numerosities presented and the manipulation to induce a distance or ratio effect, which could lead to the inconsistent results observed. The measure used to assess children's arithmetic skills also differ in the studies presented. This raises a question about the items used in the arithmetic measure and whether there are stronger relationships with some items or the type of presentation (timed/untimed) (this idea will be explored in Chapters 2, 4 and 5). In a large number of papers only simple correlations are reported (it is noted that when a large age range of children is included a control for age differences is often applied), therefore the relationships found could possibly reflect children's general ability. When associations are found the strength of relationships vary and are often quite weak. More importantly correlations do not inform us about causation, assumptions are made about the importance of the approximate number system for later symbolic arithmetic from concurrent data, and therefore longitudinal research is needed. It is important that investigations are carried out into whether these measures predict variance in later arithmetic achievement and whether these tasks are able to identify children that may go on to develop difficulties (e.g. like phonological awareness and letter knowledge are known to be important for later reading skill). The next section will therefore review the longitudinal literature.

1.5.2. Longitudinal predictors of arithmetic skill.

1.5.2.1. Symbolic comparison as a longitudinal predictor.

Whilst (in general) the cross-sectional studies show that there is a relationship between performance on symbolic comparison tasks and children's arithmetic abilities, an important topic that needs investigating is whether this ability *predicts* later arithmetic achievement. To address this De Smedt, Verschaffel and Ghesquière (2009) used a longitudinal design to examine the effect of distance on children's symbolic comparison performance and relate it to their mathematical ability one year later. In order to limit the amount of formal mathematics education they had received Time 1 of the study took place when children were beginning their formal schooling, on average 6 years of age. Children completed a symbolic comparison task and the gradient of the regression slope was used to represent the effect of distance on children's comparison speed. The mathematics measure used in this study was a curriculum based standardised achievement test (from the Flemish Student Monitoring System) and included arithmetic calculations alongside word based problems and measurement questions. It was found that children with steeper slopes gained lower mathematics achievement scores one year later (*r* = .40) and that the slope predicted 10% unique variance in children's scores over and above age, general ability and time taken to read numbers. Their findings extended the concurrent literature to show that early symbolic comparison ability is a unique predictor of later arithmetic ability (when RTs are used). Whilst children's overall accuracy on the task was related to their second grade maths performance (*r* = .38) it was not a unique predictor when age, general ability and number reading accuracy were controlled. An important limitation with the study, which the author's note, is that the autoregressor (children's early mathematics ability) was not assessed and included in the analysis. Therefore it is not known what influence the size of the distance effect had over and above children's prior mathematics skills (i.e. whether it predicts *growth* in mathematics achievement).

However, other studies have not replicated this longitudinal relationship between a measure of the distance effect and later arithmetic and mathematics achievement. Sasanguie, Van den Bussche & Reynvoet (2012) administered a symbolic comparison task (alongside other number processing measures; nonsymbolic comparison task, priming comparison task and number line estimation tasks) to 72 children who at Time 1 were kindergartners (5.6 years old), first graders (6.7 years), and second graders (7.6 years). Mathematics achievement was assessed at the first time point and one year later (Time 2) by a curriculum based standardised achievement test for mathematics (same test as used by De Smedt et al., 2009). Reaction times on the symbolic comparison task were adjusted to reflect both speed and accuracy (RT/1(1 – error)) and the effect of distance was quantified by a regression slope. While a significant correlation was found between children's mathematics achievement and mean RT when controlling for grade (*r* = -.31), no corresponding association was found with the slope (*r* = .08). Children's speed at comparing symbolic digits was found to be a significant predictor (Beta = -.61) of individual differences in mathematics achievement scores when entered alongside performance on the other number processing tasks, grade and spelling. This remained significant (Beta = -.43) when prior mathematics achievement was also included in the final step of the regression analysis.

In a similar study, Sasanguie et al. (2013) assessed both mathematics and arithmetic achievement. Mathematical achievement was assessed with the same measure used in the previous study, while arithmetic fluency was assessed with a timed test. Similar methods were also used with children assessed at the first time point ranging in age from 6 to 8 years old (first to third grade) and again at a second time one year later. However, this time the metrics used to measure symbolic comparison performance were the median RT and the distance effect (the average of the median RTs on trials with distances of 4 and 5 subtracted from the average of the median RTs on trials with distances of 1 and 2, then divided by the average of the median RTs on trials with distances of 4 and 5). Both grade and spelling ability were controlled for in the correlation analysis. Speed on the comparison task (median RT) was significantly and negatively related to achievement on both of the tests, with the strength of the association almost the same for both an untimed mathematics test and a measure assessing calculation fluency (mathematics: $r = -.37$; timed arithmetic: $r = -.35$). As before, no relationship was found with mathematics or arithmetic achievement with the effect of distance on children's comparisons. Separate regression analyses were performed for the two achievement measures with symbolic comparison RT and DE entered alongside other number processing tasks (with age and spelling ability controlled). Speed at comparing digits was found to be a significant predictor of variance in scores on both measures (untimed: Beta = -.36; timed: Beta = -.30), whereas the distance effect was not. However, when prior mathematics achievement was controlled its contribution was no longer significant. This is an interesting finding that symbolic comparison speed was not a predictor of growth in mathematics achievement. It should be noted that this finding could be partly due to a lack of power to detect the relationship; ten variables were entered into the regression analysis with a sample size of only 71. Nonetheless this finding warrants further investigation.

Using an alternate method (cluster analysis), Reeve et al. (2012) characterised children's performance on the comparison task at 6 years old as being slow, medium or fast. They found that subgroup membership at 6 years predicted arithmetic ability not only at the same time point (assessed by single digit addition) but also at 9 and a half years old (assessed by two digit addition, subtraction and multiplication) and at 11 years old (assessed by three digit subtraction, multiplication and division). The faster subgroup performed significantly better than both the medium and slow subgroups on all of the arithmetic measures at each time point, and the medium subgroup also significantly outperformed the slow subgroup.

In the majority of studies assessing symbolic comparison ability, the task is presented on a computer and the numbers range from 1 to 9. Desoete, Ceulemans, De

Weerdt and Pieters (2012) administered a different kind of task, a subtest from the TEDI-MATH Battery (Grégoire, Noël & Van Nieuwenhoven, 2004) (which also includes comparison of nonsymbolic stimuli and verbal number words). Three arithmetic measures were given which assessed untimed simple calculation skill, complex calculation ability (problems often in word problem format) and (timed) fact retrieval. The sample size was very large and included 315 typically developing children, 64 children who were classified as low achieving and 16 who were classified as having mathematical difficulties. Children were assessed in kindergarten (aged 5 to 6 years old), first grade (6 to 7 years) and second grade (7 to 8 years). No associations were found between performance on the comparison task at Time 1 and arithmetic achievement at Time 2, whereas there were significant but weak associations with Time 3 arithmetic achievement (r = .21 to .36 controlling for IQ). Children's performance on the symbolic comparison task was also found to be a significant predictor of grade 2 arithmetic achievement when entered alongside performance on the other two comparison tasks (complex: Beta = .36; simple: Beta = .28); however they did not include IQ in the regression analysis. This difference in the contribution of early comparison ability to later arithmetic achievement assessed one and two years later warrants further investigation, it is possible that the relationship changes over time. Therefore more extended longitudinal studies (e.g. Reeve et al., 2012) are needed.

1.5.2.2. Nonsymbolic comparison and as a longitudinal predictor.

The first longitudinal study to investigate the Weber fraction as a possible predictor of arithmetic achievement was carried out by Halberda et al. (2008). They found that there was great variability in the size of the Weber fraction (0.12 to 0.57) of individuals in ninth grade (on average 14 years old). This measure was then found to correlate with previous arithmetic performance assessed using the TEMA-2 and the Woodcock-Johnson Revised calculation subtest (WJ-Rcalc; Woodcock & Johnson, 1989). This was the case for every year the children were assessed, from kindergarten through to sixth grade, and over and above other cognitive measures. Whilst this study was important in establishing that there is variability in ANS acuity and these individual differences were associated with arithmetic ability at a different time point, it must be noted that the study was retrospective. Therefore the authors proposed two possible explanations for the results, that variability in the acuity of the ANS is either the cause or a consequence of the variability in arithmetic achievement. Further research needed to distinguish which, by including a measure of ANS acuity earlier in development.

In order to address this point, Mazzocco, Feigenson and Halberda (2011b) followed up the youngest children involved in their earlier study that assessed ANS acuity in children and adults (see Halberda & Feigenson, 2008). By selecting the youngest children, who completed the task whilst at preschool, meant that the measure of the ANS was taken before they had received formal mathematics instruction. Seventeen children's arithmetic achievement (TEMA-3) and general ability were assessed on average two and a half years after completion of the comparison task. Children's accuracy performance (adjusted for age and display time at preschool testing) was found to predict 28% of variance in their later arithmetic achievement scores (adjusted for age and grade at follow up testing). This relationship was found to be specific to numerical skill and not just an association with cognitive skill as no predictive relationship was found with any of the general ability measures. A Weber fraction was also calculated for each individual, and for 14 children their data conformed to the psychophysical model. This Weber fraction was found to predict 21% of variance in children's later arithmetic scores although this was not significant, although the lack of significance could be a result of a lack of power due to the small sample size. It should be noted that the estimate of the size of the Weber fraction (and the model fit to the data) were not reported here. A longitudinal study is therefore needed to expand the literature including a larger sample and as in this study using a comparison task designed to estimate a Weber fraction on an individual basis.

Addressing the sample size issue, Libertus et al. (2013) investigated the relationship between ANS acuity and arithmetic longitudinally in over 100 children (more detail of this study can be found in the concurrent relationships section). They estimated the size of the Weber fraction in young children who had received little formal arithmetic instruction (average age of 4 years 2 months, range = 2 to 6 years old) and assessed arithmetic achievement about six months later (TEMA-3). Children with smaller Weber fractions were found to have higher arithmetic scores six months later ($r = -.25$). A stronger association was found with arithmetic achievement when performance on the task was quantified as percentage of comparisons correct ($r = .46$), and children who were faster at comparing the numerosities were also found to have higher arithmetic scores ($r = -.33$). Replicating their concurrent findings, they found that accuracy and reaction time on the task predicted significant variance in arithmetic scores (7% and 6% respectively) with age, vocabulary knowledge and prior arithmetic skill controlled. Neither of the estimates of ANS acuity were significant predictors of children's later vocabulary knowledge, indicating a specific relationship with later arithmetic achievement.

However, outside of this research group the results look quite different. Two studies by Sasanguie and colleagues (Sasanguie, Van den Bussche & Reynvoet, 2012; Sasanguie et al., 2013) found that performance on a nonsymbolic comparison task was not related to mathematics and arithmetic achievement assessed one year later. It could be argued that there were some differences between the design of the study reported by Sasanguie, Van den Bussche and Reynvoet (2012) and the two previously reported (Libertus et al., 2013; Mazzocco et al., 2011b), which may have lead to different findings. The numerosities presented were smaller and some were within the subitizing range (1 to 9, which matched the symbolic version also presented), these smaller numerosities may be processed differently to items outside of the subitizing range (e.g. Revkin et al., 2008). The speed with which children performed the comparisons was also the main focus of the analysis (although RTs were adjusted to reflect accuracy) rather than accuracy (which was high) or calculating a Weber fraction. Sasanguie et al. (2013) presented a nonsymbolic comparison task to children where they compared numerosities ranging from 6 to 26 to a standard of 16. This enabled them to calculate a Weber fraction for each child; however neither the Weber fraction nor mean accuracy scores were related to their later arithmetic (fluency) or mathematics achievement.

Alongside a symbolic comparison test, Desoete et al. (2012) administered a nonsymbolic task (from the TEDI-MATH Battery). Children were assessed in kindergarten (aged 5 to 6 years old), first grade (6 to 7 years) and second grade (7 to 8 years). While significant (but weak) associations were found between performance on the nonsymbolic comparison task at Time 1 and arithmetic achievement at Time 2 (simple calculation: r = .16; complex calculation: r = .16; fluency: r = .20), only one significant (but weak) association was found with Time 3 arithmetic achievement (fluency: r = .16) controlling for IQ. Children's performance on the nonsymbolic comparison task was also found to be a significant predictor of grade 1 arithmetic achievement when entered alongside performance on the other two comparison tasks but not of grade 2 arithmetic achievement (IQ was not controlled in the regression analyses). The sizes of these associations are very weak and in studies with smaller sample sizes would no doubt be classed as a lack of a relationship. With such a large sample size it would have been beneficial to include some domain general predictors (i.e. working memory, attention) to investigate whether the contribution of performance on these magnitude comparison tasks is independent of other domain-general skills.

1.5.2.3. Differences in the longitudinal relationship with arithmetic between symbolic and nonsymbolic comparison ability.

There are fewer longitudinal than concurrent studies including both a symbolic and nonsymbolic comparison task (Desoete et al., 2012; Sasanguie, Van den Bussche & Reynvoet, 2012; Sasanguie et al., 2013). Due to the differences in the nonsymbolic comparison tasks presented it makes it more difficult to compare results across the studies. Sasanguie and colleagues (Sasanguie, Van den Bussche & Reynvoet, 2012; Sasanguie et al., 2013) found that comparison speed on the symbolic comparison task was related to later mathematics achievement whereas nonsymbolic comparison ability was not (different metrics were used for nonsymbolic comparison in the two studies). While Desoete et al. (2012) found that the accuracy at comparing nonsymbolic but not symbolic stimuli was related to arithmetic achievement one year later, whereas the opposite pattern was found when predicting arithmetic achievement two years later. These findings clearly highlight the need for a systematic, longitudinal investigation of the importance of magnitude (symbolic and nonsymbolic) comparison ability to later arithmetic achievement.

1.5.2.4. Summary of the longitudinal predictors of arithmetic achievement.

When investigating whether individual differences in performance on magnitude comparison tasks explains variance in children's later arithmetic (or mathematics) achievement the findings are again more consistent with regards to the importance of symbolic comparison ability. However, there are still inconsistent findings regarding whether the effect of distance on children's symbolic comparison decision times is an important predictor. The same can be said about the Weber fraction calculated from nonsymbolic comparison accuracy scores. Halberda and his colleagues have been able to replicate the finding that children with a smaller Weber fraction gain higher arithmetic scores later, while other groups have failed to find this relationship. As with the cross-sectional literature, various measures and analyses have been used to index performance on the magnitude comparison tasks. The outcome measures used to assess arithmetic/mathematics achievement also differ across studies, which suggest another possible factor for the differing relationships. While some studies control for variables in the analyses which may explain, in part, the relationship between the tasks (i.e. age, general cognitive ability), others include too many, while some none at all. An important point to consider is whether or not to control for prior arithmetic skill (i.e. to predict *growth*), this will be discussed in Chapter 4. What is clear from the review of the existing literature is that studies are needed that include well designed comparison tasks, looking at both symbolic and nonsymbolic comparison within the same sample and comparing the different methods of indexing performance, over an extended period of time.

1.6. Non-specific Cognitive Factors

This thesis focuses on the importance of a domain-specific predictor of arithmetic achievement but domain-general skills will also play a role. These factors may also influence the relationship between children's magnitude comparison ability their arithmetic achievement. Therefore a short discussion of the potential contribution of non-specific cognitive factors will be included here.

When thinking about the skills that are needed to answer even a simple arithmetic problem one needs to remember the numbers and the operation, whilst retrieving and executing the correct strategy to solve the problem; this will therefore draw on working memory (Hulme & Snowling, 2009). Working memory is a system that allows us to temporarily store and manipulate information and the most widely cited and influential account is Baddeley & Hitch's working memory model (Baddeley, 2000; Baddeley & Hitch, 1974). They proposed a multi-component model consisting of the central executive and two specialised slave systems, and later included the episodic buffer. The central executive is thought to be responsible for functions such as co-ordinating the slave systems, inhibition, planning, and attention (Baddeley & Logie, 1999). The two slave systems deal with the temporary storage of different kinds of information: the phonological loop deals with verbal information, while the visupospatial sketchpad deals with visual and spatial information (Baddeley, 1997). Tasks that simply involve the immediate recall of information (often lists of items, i.e. digits, words, or visual patterns) are thought to be measures of short-term memory, while those that involve both processing and storing information require additional executive and attentional processes (the central executive; e.g. listening span, counting span tasks).

Working memory has been proposed to be important for everyday cognition including developmental processes such as language development and learning to read (Baddeley, 1997). It has also been proposed that working memory is important for children's arithmetic and mathematical development (e.g. Alloway & Alloway, 2007; Bull et al., 2008; Holmes and Adams, 2006). Working memory has been found to play a role in both adults and children's arithmetic as increasing working memory load decreases performance. In a variety of experiments, Hitch (1978) presented adults with multidigit addition sums. In one experiment, participants were allowed to write their answer from right to left (units, tens, hundreds, which is typical behaviour) on half of the problems and on the other half they were made to write the answer from left to right (therefore delaying the output of partial results). They found that more errors were made when the answer had to be written in the left to right order. In another experiment they reported that increasing the amount of the sum that was presented visually, decreased the number of errors that adults made. Adams and Hitch (1997) reported that children's visual addition spans were higher than their oral addition spans. These studies also suggest that working memory demands are greater when problems are presented orally, so to reduce working memory load all arithmetic tasks presented to children in the following research will be presented visually, rather than orally.

To be consistent with the literature review on the importance of magnitude comparison performance for children's arithmetic achievement, we will now turn to the literature on working memory as a predictor of arithmetic ability. In a recent meta-analysis, including 111 studies, Friso-van den Bos, van der Ven, Kroesbergen and van Luit (2013) investigated the strength of the associations between aspects of working memory and executive functions and mathematics, and also the factors that may affect these relationships. In their search terms they included working memory measures tapping the visuospatial sketchpad, the phonological loop, visuospatial updating (more commonly referred to as visuospatial working memory as the tasks involve the central executive), verbal updating (also referred to as verbal working memory), inhibition, and shifting. They also included studies with a wide range of arithmetic and mathematical outcome measures (including for example, national curriculum tests, simple arithmetic tests, word problems, and geometry). It was reported that all working memory components showed significant positive relationships with mathematics achievement, with the strength of the associations ranging from r = .27 (inhibition) to r = .38 (verbal updating). It was also found that some factors influenced the strength of the associations. For example, the type of mathematics measure, with stronger relationships often found for general mathematic tests, than for specific measures (e.g. arithmetic, counting), and age, with stronger correlations found for younger children between shifting and visuospatial sketchpad measures and mathematics. They coded for the inclusion or not of control variables (e.g. age, IQ, reading skills), however no further distinction was made due to the low frequency of occurrence. It is therefore possible that these differences between studies could influence the findings, an issue noted in the literature on magnitude comparison and arithmetic. It also appears that the importance of working memory may differ with the measure used to assess numerical skill and also the aspect of working memory (similar issues were noted when reviewing the magnitude comparison literature).

An important question to consider is whether both domain-specific (e.g. magnitude comparison ability) and domain-general skills are important for children's arithmetic achievement when considered alongside each other. Of the studies reviewed earlier in the literature review there is a mixture of whether domain-general measures of cognitive ability were assessed or not, and whether or not they were included in the correlation or regression analyses presented. Of the twenty three papers that investigated the relationship between magnitude comparison ability and arithmetic or mathematics achievement, six did not include any domain-general measures, five included estimates of general ability (verbal or nonverbal) and four assessed reading or spelling ability to represent general cognitive ability. In general these measures were used as controls when exploring the associations between magnitude comparison performance and numeracy; therefore it is not clear whether the relationships would differ if these measures were not included as controls. There were eight papers that assessed a larger range of domain-general skills, for example, IQ, working memory, attention and processing speed, a summary of some of the findings is reported below.

Durand et al. (2005) found that there was a specific relationship between digit comparison and arithmetic in a group of 7 to 10 year olds as it was not a predictor of children's reading ability. They also included estimates of verbal ability, nonverbal ability, phonological memory, search speed and phoneme deletion ability in their path analyses and found that digit comparison and verbal ability were the only significant predictors of children's arithmetic. In a younger group of children (mean age 4 years 8 months), Libertus et al. (2013) reported that both accuracy and speed measures of the ANS were significant predictors of arithmetic ability whilst controlling for attention, memory span and vocabulary knowledge. Using a different methodology, Mazzocco et al. (2011a) compared groups of adolescents with mathematical learning difficulties (MLD) to typically developing adolescents, and groups who were low achievers and high achievers. At Grade 9 it was found that even when using digit span forward and backward and rapid automatized naming (RAN) tasks as covariates the group with MLD had significantly larger Webber fractions. They then also included additional measures which have been found to be related to arithmetic (executive function, visual memory, visual perception, digit span composite, nonword decoding and RAN) which had been assessed when the adolescents were younger. A significant main effect of mathematics group remained, with the group with MLD having the largest Webber fractions, and none of the covariates accounted for a significant amount of variance. These findings have also been replicated in longitudinal studies. Halberda et al. (2008) reported that even when controlling for 16 domain-general skills including measures of IQ, executive functions, working memory and reading, performance on a nonsymbolic comparison task (at 14 years old) was still significantly correlated with earlier arithmetic skill (5 to 11 years old). However it should be noted that this is a retrospective longitudinal study and the domain-general skills were assessed during the third grade. Therefore it is not clear whether earlier magnitude comparison ability predicts individual differences in later arithmetic skill, while controlling for domain-general skills assessed at the same earlier time point. In a prospective longitudinal study, Reeve et al. (2012) found that groups who differed in symbolic comparison speed (a measure of mean RT and effect of distance) differed in their arithmetic ability but did not differ in processing speed or nonverbal reasoning skill. Taken together these studies suggest that the relationship between the approximate number system (assessed using magnitude comparison tasks) and arithmetic and mathematics achievement is not attributable to domain-general skills. It should also be noted, as has been reported earlier, the evidence for a relationship between comparison performance and arithmetic achievement is mixed, and often a relationship has not been found, however, it does not appear that this difference is due to the inclusion (or not) of domain-general measures.

It should be noted that more recently studies that have investigated inhibitory control have found different results. Both Fuhs and McNeil (2013) and Gilmore et al. (2013) found that the relationship between accuracy on nonsymbolic comparison tasks and arithmetic was driven by children's performance on incongruent trials, this is where the surface area of the stimuli conflicts with numerosity (i.e. as numerosity increases the surface area decreases). Both studies also reported that in regression analyses when a measure of inhibitory control skill was entered before performance on the nonsymbolic comparison task it was no longer a significant predictor of arithmetic achievement. Fuhs and McNeil (2013) proposed that inhibitory control may drive the relationship found between magnitude comparison performance and numeracy in young children and Gilmore et al. (2013) proposed that it is the inhibition element in these nonsymbolic comparison tasks that is related to arithmetic rather than the precision of representations within the ANS. The nonsymbolic comparison tasks used in this research will therefore aim to control for perceptual variables.

1.7. Summary and Research Aims

It has been proposed that even young children have a system for representing numerical magnitude nonverbally (approximate number system), that the symbolic number system may map onto this, and therefore forms the basis for which formal arithmetic knowledge may build upon. Measures which tap the acuity of this system have been designed - magnitude comparison tasks, these simply involve choosing the larger numerosity and those that include nonsymbolic rather symbolic stimuli can be presented to young children. Multiple indices have been used to quantify performance (e.g. accuracy, reaction time, distance/ratio effect, Weber fraction). Over development, performance on these tasks improves and the effect of the distance between numerosities on comparison ability decreases, as does the size of the Weber fraction. This has been taken to suggest that the acuity of representations within the approximate system become more precise (whether this is due to less compression or less variability is unknown and also difficult to entangle).

While the change in the distance (and ratio) effect and size of the Weber fraction has been investigated in different age groups the majority of these studies are cross-sectional, i.e. include different children. What is missing from the existing literature is the development of this in the same children over the initial period of formal schooling, i.e. longitudinal data. As many different ways to investigate the effect of distance and to estimate the acuity of the ANS have been used with differing findings, it will be important to address this issue within the construction of the tasks and the analysis of the data.

The importance of this approximate number system on arithmetic development has therefore become an emerging area of interest. There is consensus that some aspect of symbolic comparison ability is related to individual differences in arithmetic achievement in different age groups of children both concurrently (e.g. Durand et al., 2005; Holloway & Ansari, 2009; Mundy & Gilmore, 2009) and longitudinally (e.g. De Smedt et al., 2009; Sasanguie, Van den Bussche & Reynvoet, 2012; Sasanguie et al., 2013). Whether this is accuracy, speed, or the effect of distance on the task has varied along with the arithmetic or mathematics measure used to represent this ability.

The findings investigating the relationship between measures of nonsymbolic comparison and arithmetic are less clear. Halberda and colleagues (Halberda et al., 2008; Libertus et al., 2011; Libertus et al., 2013; Mazzocco et al., 2011b) consistently find a relationship between the two skills both concurrently and longitudinally, in different age groups, using different presentations of the task, and when also controlling for other cognitive skills. Whereas less consistent results are found by other groups (e.g. Holloway & Ansari, 2009; Mundy & Gilmore, 2009; Sasanguie et al., 2013). There are many differences between the methods used which could be possibilities for these inconsistencies, for example the way the comparison task is presented, the number of items being compared, the method of analysing the data, and the arithmetic measure used to represent numerical skill. This warrants further investigation within the same sample of children, for example by completing nonsymbolic comparison tasks with these different manipulations and by calculating multiple indices of performance.

A more systematic approach should therefore be taken when investigating the relationship between both symbolic and nonsymbolic comparison and arithmetic achievement, to enable further understanding of this association. Further research is needed with a longitudinal design to investigate the relationship of magnitude comparison ability with arithmetic to expand the existing literature. Whilst studies incorporating a longitudinal design have been conducted it will be important to include measures of both arithmetic and magnitude comparison at all time points, this will enable the investigation of the bidirectional relationship between the two. Children's early arithmetic skill should also be controlled in the analysis which means that it is the *growth* in arithmetic knowledge that is being investigated. When prior arithmetic skill is not controlled any relationships found between magnitude comparison and later arithmetic may be reflecting the association between arithmetic ability at the two time points, rather than a specific relationship between the two different measures (magnitude comparison and arithmetic). At an early time point arithmetic ability and magnitude comparison ability may be sharing variance, therefore by controlling for prior differences in arithmetic skill you are removing the variance due to early arithmetic ability.

The differing findings of previous research may be due to the different arithmetic and mathematics measures used. Many of the measures include items that could be described as assessing basic understanding of number rather than calculation abilities and some of the items may overlap with the processes involved in the magnitude comparison tasks (e.g. identifying larger numbers, ordering numbers). Therefore it is important that future research also includes measures that avoid this possible overlap. Another suggestion would be to include both timed and untimed measures to investigate whether there is a specific relationship with arithmetic fluency (speed at which simple problems are solved or retrieved from memory) or with arithmetic ability in general (i.e. also more difficult calculation problems). In addition the relationship between magnitude comparison and arithmetic may differ with different operations (addition, subtraction, multiplication). Therefore to extend this research area an examination of the relationships between magnitude comparison and these different numerical operations assessed separately must be carried out as well as with overall calculation ability.

An additional point to note is that in some studies the age range of the children spans two or three years or grades. To help clarify the relationships a more concentrated age range should be used, with the measures of magnitude comparison administered as early as possible before children have received a large amount of formal schooling in numeracy.

As highlighted earlier, there are other domains that will be important for children's arithmetic development (i.e. working memory, counting, attention), as well as environmental influences, for example teaching and social economic status. To include every possible predictor of children's arithmetic development is beyond the scope of this thesis; however, children's verbal and nonverbal ability will be estimated to account for their general cognitive ability, this will allow for the investigation of whether any relationships found are due to individual differences in this ability.

The main aims of this thesis are therefore:

- To assess the relationship between symbolic and nonsymbolic comparison tasks and whether they are tapping the same underlying construct.
- To investigate performance on these magnitude comparison tasks over development, within the same sample of children (i.e. longitudinally).
- To explore the concurrent relationships between numerical processing and arithmetic achievement, while controlling for children's general cognitive ability, in a large representative sample.
- To extend this longitudinally by also exploring the predictive value of these measures to later arithmetic achievement over an extended period of time (one and two years).
- In order to address some of the previous aims a large number of children will need to be recruited and the most efficient way to complete this will be to use a group testing design. Therefore in order to verify the findings, a subgroup of children will

also need to complete individually presented measures, more akin to those frequently used in the literature.

Chapter 2.

Magnitude Comparison as a Concurrent Predictor of Children's Arithmetic Achievement

2.1. Introduction

The overall aim of this first study was to explore the predictors of children's arithmetic ability with a specific focus on children's magnitude comparison ability. Magnitude comparison tasks, both symbolic and nonsymbolic, have been used to investigate the proposal of an internal system for representing number in humans (and some nonhuman animals). These tasks have typically involved comparing Arabic numerals or nonsymbolic stimuli such as dots or squares. The tasks are very simple and involve deciding which of two presented items is the largest in terms of its magnitude. However, most studies presented in the literature raise a question about the nature of the comparison tasks and their underlying cognitive processes. To my knowledge no study has fully investigated commonalities and differences between different comparison tasks. To investigate this, three different types of comparison task were created; digit (symbolic), nonsymbolic and letter. Digit comparison involves comparing single-digit Arabic numerals; therefore children will need to understand the magnitude that is represented by the symbol. Nonsymbolic comparison involves comparing random patterns of small squares; these included smaller (5 to 13) and larger (20 to 40) numerosities. The final type of comparison task involved comparing letters of the alphabet and indicating which came later in the alphabet sequence. This task therefore acted as a control for the comparison process.

The first aim of this study was to assess the underlying latent factors that different comparison tasks may have in common. First, do all comparison tasks (digit, nonsymbolic and the letter control task) represent one latent factor? It is possible that all three types of comparison task tap a general comparison construct which is reflected in the ability to compare two items and choose one of them. Second, the comparison tasks could form two latent factors: a magnitude (digit and nonsymbolic) comparison factor and a separate letter comparison factor. The magnitude factor would be specific for magnitude representations, i.e. for comparisons about numerical size. This construct would be involved in symbolic (digit) and nonsymbolic magnitude comparison but not in letter comparison. Another possible latent two factor structure could emerge with a symbol comparison construct (digit and letter) and a separate nonsymbolic factor. Digits and letters are both learned symbols, and performance on the tasks might depend on the ability to learn and abstract information from symbols. In addition digits and letters both have an ordinal structure and the comparison of them might call upon sequence or ordinal representations. They are also high frequency items that are learned by heart, which is in contrast to the nonsymbolic items. Finally, a three factor solution may emerge as each task involves comparing different items (digits, nonsymbolic stimuli, and letters), and each type of comparison task may rely on a different mechanism and way to complete the task.

To allow for the investigation of the underlying factor structure multiple comparison tasks that were presented and scored in the same way were created for the present study. The tasks measured both accuracy and speed with which comparisons were made (items correct within 30 seconds); because both measures have been used in various studies (see Chapter 1).

With regards to the design of the comparison tasks, the effects of distance and ratio that are typically seen in the literature were investigated. The distance effect is a highly replicable effect; items that are numerically close to each other are more difficult to compare which is reflected in lower accuracy rates and longer comparison times than when comparing two items with a larger numerical distance between them. This is true when comparing symbolic digits (e.g. Holloway & Ansari, 2009; Moyer and Landauer, 1967), nonsymbolic stimuli (e.g. Holloway & Ansari, 2009; Mussolin et al., 2010), and also when indicating which of two letters is later in the alphabet (e.g. Jou & Aldridge, 1999; Parkman, 1971). Another way to investigate the effect of distance, which takes general problem size into account, is to manipulate the ratio of the numerosities being compared, for example individuals have been found to be more accurate and faster at choosing the larger numerosity when the ratio between them is 1:2 than when it is 7:8 (Halberda et al. 2008).

The study used a novel methodology (group presentation of the tasks) so multiple comparison tasks were created. The distance and ratio between the items presented were manipulated and presented as separate tasks. Digits (1 to 9), letters (a to i) and nonsymbolic small range items (5 to 13) were of a close distance (difference of 1 or 2) and far distance (difference of 5, 6, and 7), while nonsymbolic large range items (20 to 40) varied by the ratios 3:4, 5:6 and 7:8. Each of these manipulations was presented separately. The measures created for this study encompassed both accuracy and speed (number of items correct in 30 seconds) so it was hypothesized that a distance effect would be observed with children gaining higher scores on the far comparison task than on the close comparison task, across each item type. With regard to the ratio manipulation it was hypothesized that an effect of ratio would be observed, with scores decreasing as the ratio between the magnitudes decreased (number of items correct in 30 seconds for 3:4>5:6>7:8). If the intended effects were observed then it could be assumed that this novel methodology was successful and children were processing these group presented paper and pencil comparison measures in the same way as they would if they were presented in a one to one setting on a computer.

It has been proposed that these magnitude comparison tasks are a measure of an individual's internal representation of number (e.g. Moyer & Landauer, 1967) and this is believed to be the foundation for more complex arithmetic skill to build upon (e.g. Dehaene, 1997). If this indeed is the case then a relationship would be expected with formal arithmetic achievement. A second aim of this chapter was therefore to investigate the relationship between these comparison tasks and arithmetic.

As reviewed in Chapter 1, in general, there is consensus that children's performance on symbolic comparison tasks is related to their arithmetic achievement (see Table 2.1). This relationship is more consistent when speed, rather than accuracy, is analysed. The less consistent relationship between accuracy on the task and arithmetic ability is possibly due to high accuracy performance on the task which results in a lack of variance in scores. In the current chapter a measure that taps both accuracy and speed was used, whereas in Chapter 6 the predictive power of speed and accuracy performance will be investigated separately.

Table 2.1.

Table of published studies investigating the concurrent relationship between arithmetic and symbolic comparison

Paper	Age (Y;M where possible)	N	Number range	Arithmetic/mathematics task	Classification[a]	Correlation with arithmetic (r)	
						Accuracy	RTs
Budgen & Ansari (2011)	7;04 (1st & 2nd graders)	119	1 to 9	Woodcock-Johnson III Tests of Achievement (WJ-III) - Calculation - Fluency	Arithmetic		Controlling for age: r = -.21* to -.42** (slope intercept, ratio effect) No relationship between ratio effect and calculation (r = -.15)
De Smedt & Gilmore (2011) - TD, LA, MLD study - Belgium	6;08	82		Curriculum based standardised achievement test for mathematics, Math up to 10[b]	Mathematics	Overall accuracy was high – no differences between the 3 groups	Children with MLD were significantly slower than TD children, the LA children did not differ from the other two groups
Durand, Hulme, Larkin & Snowling (2005) - UK	7 to 10	162	1 - 9	1 minute addition 1 minute subtraction WOND: Numerical operations	Arithmetic	Items correct in 30 seconds r = .37** - .45** (controlling for age)	
Holloway & Ansari (2009)	6, 7, 8 yrs	N = 87 (6 = 29 7 = 31 8 = 27)	1 to 9	WJ-III: - Fluency - Calculation - Composite	Arithmetic	NDE and Mean accuracy – no significant correlations	NDE r = .19 (ns) to .34** Mean RT significant correlation with fluency only, r = -.37**
Landerl & Kölle (2009)	7;06 to 11;06	N = 313 TD = 262 Dyscalculic = 51	1 to 7	Heidelberger Rechentest (HRT)[C] - Addition - Subtraction - Multiplication	Arithmetic		Slope Addition r = .13* Subtraction r = .11 ($p < .06$) Multiplication r = .03 (ns)

Mundy & Gilmore (2009). - UK	7;04 (Year 2)	33	1 - 9	Created their own	Arithmetic	Overall accuracy: composite = .52** calculation = .53**	NDE composite = -.52** calculation = -.54**
Sasanguie, De Smedt, Defever & Reynvoet (2012) - Belgium	5, 6, 7, 11 yr olds	124	1 to 9	Curriculum based standardised achievement test for mathematics[b] (Kindergarten children had a different test)	Mathematics		Slope r = .22*, RT r = -.24** (adjusted RT reflects speed & accuracy)
Vanbinst, Ghesquiere & De Smedt (2012) - Belgium	8;10	49	1 to 9	Curriculum based standardised achievement test for mathematics[b]	Mathematics		Slope r = .41**

Note. TD = typically developing, MLD = mathematical learning disability, LA = low mathematics achievement, HA = high achievement in mathematics. WF = Weber fraction. [a] to aid interpretation of the results the measures used in the studies were classified as either mathematics or arithmetic. Arithmetic is a branch of mathematics so the term arithmetic will be used when the task assesses basic number skills (e.g. identifying and writing numbers and counting) and basic calculations involving addition, subtraction, multiplication and division. The term mathematics will be used when a task also includes problems outside of basic calculations, for example geometry, algebra. [b] this measure included number knowledge, understanding of operations, arithmetic, word problems, measurement, and geometry. [c] HRT; Haffner, Baro, Parzer & Resch (2005).
* $p < .05$, ** $p < .01$

In contrast the evidence for a relationship between arithmetic achievement and performance on nonsymbolic comparison tasks is less clear (see Table 2.2). This could be due to multiple reasons: the diverse comparison tasks used by the different studies, whether performance on a symbolic version is also considered, and also the way performance on the task is conceptualised. The symbolic comparison tasks used are similar in presentation, for example they involve the comparison of digits 1 to 9 (see Table 2.1), whereas nonsymbolic comparison tasks often vary in the number of items compared (see Table 2.2). For example, Holloway and Ansari (2009) presented 1 to 9 items (to match their symbolic comparison task) and found no relationship with arithmetic, whereas Inglis et al. (2011) found a significant relationship between comparison performance and arithmetic when children compared 5 to 22 items. Therefore this study included both a smaller and larger range of items to be compared. Another inconsistency in the literature is whether performance on a symbolic comparison task is controlled for or not. Mundy & Gilmore (2009) found that performance on a nonsymbolic comparison task was significantly related to arithmetic achievement but was not a predictor of arithmetic achievement when performance on a symbolic version was controlled. This highlights another reason for including both symbolic and nonsymbolic comparison tasks in the present study and to examine their relative contribution.

Table 2.2.

Table of published studies investigating the concurrent relationship between arithmetic and nonsymbolic comparison

Paper	Age	N	Number range	Arithmetic/mathematics task	Classification[a]	Correlation with arithmetic (r)	
						Accuracy	RTs
Bonny & Lourenco (2013)	3, 4, 5 yrs	74 3 = 24 4 = 25 5 = 25	4 to 12 with a fixed reference of 8	TEMA-3[b]	Arithmetic	Use predicted accuracy at an untested ratio. Across age groups r = .39** 3 year olds r = .38 (p = .068) 4 year olds r = .40* 5 year olds r = .13 (NS)	
De Smedt & Gilmore (2011) - TD, LA, MLD study - Belgium	6;08	82	1 to 9	Curriculum based standardised achievement test for mathematics[c]	Mathematics	No group differences	No group differences (RT or DE)
Fuhs & McNeil (2013)	4;07 (3;05 to 5;11)	103	1 to 30 (not computerised) Ratios:1:4, 1:2, 2:3, 3:4, 4:5, 5:6, 6:7, 7:8, 8:9, 9:10	TEMA-3[b]	Arithmetic	No significant relationship	
Gilmore et al. (2013)	4:08 to 11;11	80	5 to 22 Ratios = 0.5, 0.6, 0.7, 0.8	WJ-III: Calculation	Arithmetic	Overall accuracy: r = .57** (incongruent trials r = .55**, congruent trials r = .03, ns)	
Holloway & Ansari (2009)	6, 7, 8 yrs	N = 87 6 = 29 7 = 31 8 = 27	1to 9	WJ III: Fluency Calculation Composite	Arithmetic	No significant relationships with any of the arithmetic measures (using NDE or % correct)	Only 1 significant correlation, Mean RT & Fluency r = -.21*

Inglis, Attridge, Batchelor & Gilmore (2011)	Children: 8;05 (7;07 to 9;05)	39	5 to 22 Ratios= 0.5, 0.6, 0.7, 0.8	WJ-III: Calculation	Arithmetic	Weber fraction (WF): r = -.55* (after controlling for WASI matrix reasoning score and age, n=24)	
	Adults: 23 (18 to 48)	101	9 to 70 Ratios = 0.625, 0.714, 0.833	WJ-III: Full battery	Arithmetic & Mathematics	No significant relationships (WF)	
Libertus, Feigenson & Halberda (2011)	4;02 (33 to 73 months)	N = 174	4 to 15	TEMA-3[b]	Arithmetic	% correct r = .42** WF r = .26**	Overall RT r = .28**
Libertus, Feigenson & Halberda (2013) - Also longitudinal	Time 1 4;02 Time 2 4;09	Varies >100	4 to 15	TEMA-3[b]	Arithmetic	Time 1: % correct r = .44** WF r = -.26** Time 2: correct r = .52** WF r = -.42**	Time 1: Mean RT r = -.28** Time 2: Mean RT r = -.36**
Libertus, Odic & Halberda (2012)	Adults	120	? Ratios: 2.0, 1.50, 1.33, 1.25, 1.16, 1.1, 1.09, 1.07, 1.05	Scholastic Aptitude Test (SAT) scores	?	WF r = -.22*	
	Adults	61	Ratios: 2.47, 1.43, 1.28, 1.18	Scholastic Aptitude Test (SAT) scores	?	WF r = -.31* (although session 1 r = -.19, session 2 r = -.31*)	
Lyons & Beilock (2011)	Adults 20;05	54	1 to 9	Mental arithmetic problems. Adapted from the Kit of Factor-Referenced Cognitive Tests	Arithmetic	WF r = -.34*	

Mazocco, Feigenson & Halberda (2011a) - TD, MLD, LA, HA study - Concurrent & longitudinal (retrospective)	14;10 (14;02 to 15;11)	71	5 to 16 Ratios = 1:2, 3:4, 5:6, 7:8	TEMA-2[b] WJR-Calc (used to determine the mathematics achievement groups)	Arithmetic	Children with MLD had a significantly higher Weber fraction scores than the other 3 groups. No other group differences.	
Mundy & Gilmore (2009) - UK	7;04 (Year 2)	33	1 to 9	Created their own	Arithmetic	r = .35	No significant relationship (DE)
Piazza et al. (2010) - Children with Dyscalculia	10;08 (8;03 to 12;07)		12 to 20 & 24 to 40	Regrouped dyscalculia battery - Semantic - Complex calculation - Transcoding - Arithmetic fact retrieval	Mathematics	Only 1 significant relationship: WF with the scores on the semantic tasks, r = .41*	
Price, Palmer, Battista & Ansari (2012) 3 NS tasks - intermixed - paired - sequential	Adults	36	6 to 40 Ratios: .25, .33, .50, .66, .75, .90	WJ-III: Math Fluency	Arithmetic	WF: no significant relationships with any presentation type (intermixed r = -.25, sequential r = -.10, paired r = -.28)	Slope: no significant relationships with any presentation type (don't present values)
Sasanguie, De Smedt, Defever & Reynvoet (2012) - Belgium	5, 6, 7, 11 yr olds	124	1 to 9	Curriculum based standardised achievement test for mathematics[c] (Kindergarten children had a different test)	Mathematics		No relationship whether using slope (r = -.16) or adjusted RTs to reflect both speed and accuracy (r = .08)

Vanbinst, Ghesquiere & De Smedt (2012) - Belgium	8;10	49	1 to 9	Curriculum-based standardised achievement test for mathematics[c]	Mathematics	No significant relationship (DE r = .03)

Note. TD = typically developing, MLD = mathematical learning disability, LA = low mathematics achievement, HA = high achievement in mathematics. WF = Weber fraction. [a] to aid interpretation of the results the measures used in the studies were classified as either mathematics or arithmetic. Arithmetic is a branch of mathematics so the term arithmetic will be used when the task assesses basic number skills (e.g. identifying and writing numbers and counting) and basic calculations involving addition, subtraction, multiplication and division. The term mathematics will be used when a task also includes problems outside of basic calculations, for example geometry, algebra. [b] TEMA: Test of Early Mathematics Ability - this measure included numbering skills (e.g. counting number of objects), number comparison, numeral literacy (e.g. reading Arabic numerals), mastery of number facts (e.g. retrieving multiplication facts), calculation skills (e.g. written addition & subtraction problems), number concepts (e.g. answering how many 10s in 100). [c] this measure included number knowledge, understanding of operations, arithmetic, word problems, measurement, and geometry. Where there is a question mark, not enough detail was provided in the paper.

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

The majority of the studies reviewed do not control for children's knowledge and familiarity with the symbols (Arabic numerals) used in the symbolic comparison task and the arithmetic measure. This is important as the relationship between symbolic comparison and arithmetic achievement could reflect this rather than a relationship with the underlying representations of magnitude (Mazzocco & Thompson, 2005). The symbolic comparison tasks require children to understand the magnitude that a symbol represents, whereas comparison of nonsymbolic items does not. This could explain the more consistent evidence for a relationship between arithmetic achievement and symbolic comparison than with nonsymbolic comparison. A measure of children's ability to write and identify numbers was therefore included in this study.

A weakness of many of the studies investigating magnitude comparison is that they have often not included a task controlling for the comparison process itself (e.g. van Opstal, Gevers, De Moor & Verguts, 2008; van Opstal & Verguts, 2011). A control task was therefore included in this study to allow for the investigation of whether (if any) relationships found between arithmetic ability and comparison performance reflect an association simply with the comparison of two stimuli or a relationship with the comparison of two items that include magnitude information (digits and nonsymbolic tasks only). A letter comparison task was therefore created, this also acts as a control for the symbolic comparison task as it also involves the comparison of items that have an ordinal structure and both the counting sequence and the alphabet are over learned stings of items.

Very few previous studies have controlled for children's general ability in the analysis. Durand et al. (2005) found that children's verbal ability was a significant predictor of arithmetic achievement alongside digit comparison performance. Therefore it is important to investigate whether magnitude comparison performance is significantly related to arithmetic achievement independently of children's general cognitive ability. This study therefore included measures of both children's verbal (vocabulary) and nonverbal ability.

Age is a factor that is often controlled for to account for any maturational differences between the children. For this reason age will also be entered into the analyses even though children recruited for this study were all in the same school year group (Year 1) and should have received the same amount of schooling.

Although some studies have found a relationship between magnitude comparison performance and arithmetic, on closer inspection a number of these correlations are relatively weak and are not consistently found between the different ways to analyse comparison performance (i.e. accuracy and speed) and arithmetic (see Table 2.1 and Table 2.2). For example Budgen and Ansari (2011) found associations between calculation scores and different measures of symbolic comparison performance ranging from $r = -.15$ (ns) to $-.22$ ($p < .05$). While Sasanguie, De Smedt et al. (2012) found significant correlations between mathematics and symbolic comparison performance, they were however weak in strength (adjusted RT: $r = .24$; slope: $r = .22$). This poses the question about the predictive power of these measures.

A further question posed from the literature regards the actual arithmetic measure used to assess arithmetic achievement. The relationship between magnitude comparison and arithmetic might differ due to the type of measure used. Numerous studies have used standardised tests for example the Woodcock-Johnson III Tests of Achievement (WJ-III), which contain both untimed and speeded subtests (e.g. Holloway & Ansari, 2009), or used standardised untimed achievement tests based on the school curriculum (e.g. De Smedt et al., 2009). Other studies have created their own arithmetic measure (e.g. Mundy & Gilmore, 2009). Budgen and Ansari (2011) assessed arithmetic using the WJ-III and found stronger relationships with the speeded subtest (fluency) than with the untimed measure. Therefore the current study presented children with both a standardised (untimed) arithmetic measure and speeded calculation tasks that assessed children's ability to solve both addition and subtraction problems (arithmetic facts) separately. Therefore children's arithmetic rather than mathematical ability was assessed. The measures chosen presented the sums in their basic equation form, so were all written arithmetic tests. This was favoured over a measure that presented the items as word problems, which would place a heavier influence on children's reading skill, or presenting the problems verbally, which may place a greater demand on working memory capacity.

In addition to a careful design of basic numerical tasks, sophisticated data analysis techniques were used, taking advantage of the large sample. A large number of children were recruited and the data were analysed with Structural Equation Modeling. There are many advantages to using Structural Equation Modeling (SEM). First, SEM is a priori; requiring the researcher to construct a model which is then tested for its compatibility with the data. Second, SEM can include both observed and latent variables. Third, it enables the researcher to build and test the significance of complex models at both the micro- and macro-level. Fourth, measurement error can be taken into account in the model, meaning that relationships between variables can be modelled with perfect reliability. Finally, because SEM requires a relatively large sample size, the overall findings are more representative (see Bryne, 2012; Kline, 2011; Werner & Schermelleh-Engel, 2009).

In summary, the current study aimed to address two main questions. The first was to assess the underlying latent factors that different comparison tasks may have in common. Therefore children at the start of formal schooling were presented with multiple comparison tasks, that included symbolic, nonsymbolic and letter comparison measures so that the underlying latent constructs of the tasks could be investigated. The second aim was to investigate the relationship between the comparison tasks and arithmetic achievement. By including the letter comparison task as a control, this allowed for the investigation of whether magnitude comparison had a specific relationship with arithmetic, or whether it is just the process of comparing two stimuli (general comparison performance) that is related to arithmetic achievement. As mentioned previously very few studies control for other variables that may have influence on the relationship. Therefore the age of the children was taken into account in the subsequent analysis, alongside their general cognitive ability. For children to be able to complete both the arithmetic tests and the symbolic comparison task they must be familiar with digits and understand what they represent, therefore children's ability to correctly identify and write Arabic numerals was also assessed. Finally, data was analysed using sophisticated techniques, i.e. SEM, which required a large sample of children to be recruited.

2.2. Method

2.2.1. Participants

One hundred and seventy five children were recruited from four primary schools in the North and West Yorkshire regions. Children range from socially disadvantaged to socially and economically advantaged backgrounds, though all four schools have lower than average numbers of children entitled to free school meals. Three of the schools are larger than average whilst one is smaller than average (all information retrieved from Ofsted reports: Ofsted, n.d.). Children were tested from eight classes overall, three of these were joint Year 1 and 2 classes. All children in Year 1 were invited to take part in the study and therefore ranged in age from 5 years 8 months to 6 years 9 months. Consent to carry out the study was gained from the schools and Headteachers. Data were excluded on an individual case basis where children were deemed not able to access the presentation of the tasks. This resulted in data being removed from one child where English was an additional language and one child who had special educational needs. The final sample included 173 children (97 males) with a mean age of 74.69 months (SD = 3.43).

2.2.2. Assessment Battery

Children were assessed on the following measures.

2.2.2.1. Nonverbal ability.

Nonverbal ability was estimated using a matrix reasoning task. Sets A, B, and C of the Raven's Standard Progressive Matrices – Plus (Raven's SPM-Plus; Raven, 1998) were chosen. Children were given an incomplete matrix and asked to indicate from a choice of six or eight missing pieces which would complete the matrix. Children indicated their chosen answer directly on the matrices booklets to simplify the task. Children were awarded one point per correct response with a maximum possible score of 34.

2.2.2.2. Vocabulary.

Vocabulary knowledge was estimated using 36 items taken from sets five to seven of the British Picture Vocabulary Scale 3rd Edition (BPVS – III; Dunn, Dunn & Styles, 2010) with the publisher's permission. Children indicated which of four pictures best matched the spoken target word. The pictures were presented using a whiteboard and the children indicated their chosen answer by ticking a black and white version in their answer booklet. The presentation of this task was split over two sessions. Children were awarded one point for each correct item (maximum score of 36).

2.2.2.3. Arithmetic.

Arithmetic skill was assessed using a standardised test, the Numerical Operations subtest of the Wechsler Individual Achievement Test-Second UK Edition (WIAT-II^{UK}; Wechsler, 2005) and two speeded arithmetic tasks: one-minute addition and one-minute subtraction (Westwood, Harris-Hughes, Lucas, Nolan & Scrymgeour, 1974).

2.2.2.3.1. WIAT-II: Numerical Operations subtest.

The Numerical Operations subtest was adapted for group use by taking the first 16 items. Item 6 was removed as it requires a verbal response. Following the administration guide children were allowed to complete the task in their own time. The subtest begins with identifying and writing Arabic numerals and progresses to simple addition, subtraction and multiplication problems (see Appendix 1 for the full list of items). To gain a correct score (one per item) the Arabic numeral had to be written correctly, numerals that were reversed/written backwards (i.e. 6 written as ∂) were scored as incorrect. The maximum achievable score was 15.

2.2.2.3.2. One minute tasks.

To asses speeded arithmetic skill (calculation fluency) all children completed a one minute addition task; children were given one minute to complete as many simple written addition problems (e.g., 2+1, 3+4) as possible (see Appendix 2). A subset of children (n = 79) completed the subtraction version of this task (e.g., 5-1, 6-2) (see Appendix 3). Problems were presented as a list of 30 items; therefore a maximum score of 30 was possible for both tasks (items correct in 60 seconds). Responses where individual numerals were written backwards were scored as correct (i.e. 3 as Ɛ, or 15 as 1ƨ) but when the order of a multidigit number was written incorrectly (i.e. 15 as 51), the answer was scored as incorrect.

2.2.2.4. Number writing.

To assess children’s ability to write Arabic numerals correctly a task was created. Children were asked to transcribe the following Arabic numerals: 12, 19, 20, 37, 63, 100, and 152 which were presented verbally. The numeral must have been written correctly to be scored as correct, reversed numerals or numbers in the incorrect order were scored as incorrect. The maximum score was seven.

2.2.2.5. Number identification.

To assess children’s ability to match Arabic numerals (symbolic form) with their verbal label a number identification task was created (see Appendix 4). The experimenter read aloud a number and the children were asked to choose the correct written format from a choice of four or five presented. Distractor items were chosen to include common errors that young children make. For example if 163 was the target item the children were presented with the choices of: 13, 10063, 136, 16, **163**. Children were presented with the following items: 6, 14, 28, 52, 76, 163, 235, and 427. The maximum score on this task was eight.

2.2.2.6. Comparison tasks.

Multiple comparison tasks were created for the study; they were based on those typically used in the literature (e.g. Moyer & Landauer, 1967) and were adapted for large group testing so they could be presented as paper and pencil measures. Fourteen comparison tasks were constructed in total but only those relevant for the analysis are presented here. Each comparison task was presented individually within one of two A5 booklets. For many of the tasks there were two similar versions (e.g. digit close and digit far); these were separated and presented in the different booklets that were given in different sessions on different days. The comparison pairs were presented in a pseudo random order with the restriction that the same item never appeared directly underneath itself. Six comparison pairs were presented per page. For each comparison task children were given 30 seconds to complete as many comparisons as they could. For every correct comparison children were awarded one point with the maximum score on each version being 36. The full list of comparison tasks are detailed in Appendix 5 and an example of how the comparison tasks were presented to the children is in Appendix 6.

2.2.2.6.1. Symbolic digit comparison.

The items ranged from 1 to 9 and the pairs of digits were presented in both orders (5 and 3, 3 and 5). To allow for the investigation of the effect that distance had on comparison (numerical distance effect), two versions of the task were created: close (comparison items with a numerical distance of 1 and 2) and far (numerical distance of 5, 6, and 7). The two versions of the task were matched on the total problem size of the two digits being compared, for example the close comparison of 5 and 3 was matched to the far comparison of 7 and 1. The pairs of Arabic digits were presented within boxes ($2cm^2$), one on the left and one on the right (see Figure 2.1 for an example). All digits were presented in Calibri font size 48.

2.2.2.6.2. Letter comparison.

A letter comparison task presented in the same format as the digit comparison task was constructed (see Figure 2.1). The letters a to i were used with the letter 'a' replacing the digit 1, 'b' replaced 2 and so on. All letters were presented in lower case format in Comic Sans MS font size 48. This font was chosen so that the letter 'a' would be presented in a way that the children would recognise i.e. ɑ. The letter comparison tasks also included two versions: close and far.

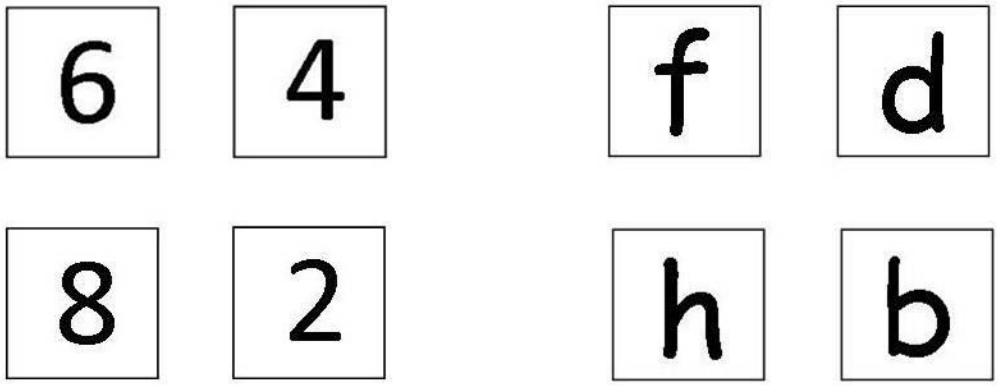


Figure 2.1. Example of digit and letter comparison stimuli at both the close (above) and far (below) comparison distances.

2.2.2.6.3. Nonsymbolic comparison.

All stimuli consisted of arrays of black squares within a 2.5cm^2 black border on a white background and were presented in pairs, one on the left and one on the right. There were two versions of the nonsymbolic task, one to investigate the *distance effect* and a second to investigate the *effect of ratio*. These tasks varied in the size of the numerosities presented. A second manipulation was made to these tasks with a further two versions of each being constructed. In one version the squares used in each stimulus pairing were the same size (SS) meaning that larger numerosities had more squares and therefore a larger surface area of black. In the other, the two stimuli pairings were matched for total surface area (SA), therefore smaller numerosities had larger squares but the same surface area of black. By including stimuli that was half of the time the same size and the other half controlled for total surface area it was hoped that children would discriminate based on numerosity rather than nonnumerical cues. All stimuli were created using Microsoft Paint and ArcSoft PhotoStudio 5.5 and more detail on the construction of the same size and surface area matched stimuli can be found in Appendix 7.

2.2.2.6.3.1. Effect of distance.

This task was matched to the format of the digit and letter comparison tasks but to avoid the subitizing range the numerosities presented ranged from 5 to 13 (see Figure 2.2). The numerosity 5 replaced the digit 1, 6 replaced 2, and so on. Again there were two versions: close (numerical distance of 1 or 2) and far (numerical distance of 5, 6 or 7). As in the symbolic comparison tasks the two versions were matched on the total problem size of the two numerosities being compared, for example the close comparison of 10 and 8 was matched to the far comparison of 12 and 6.

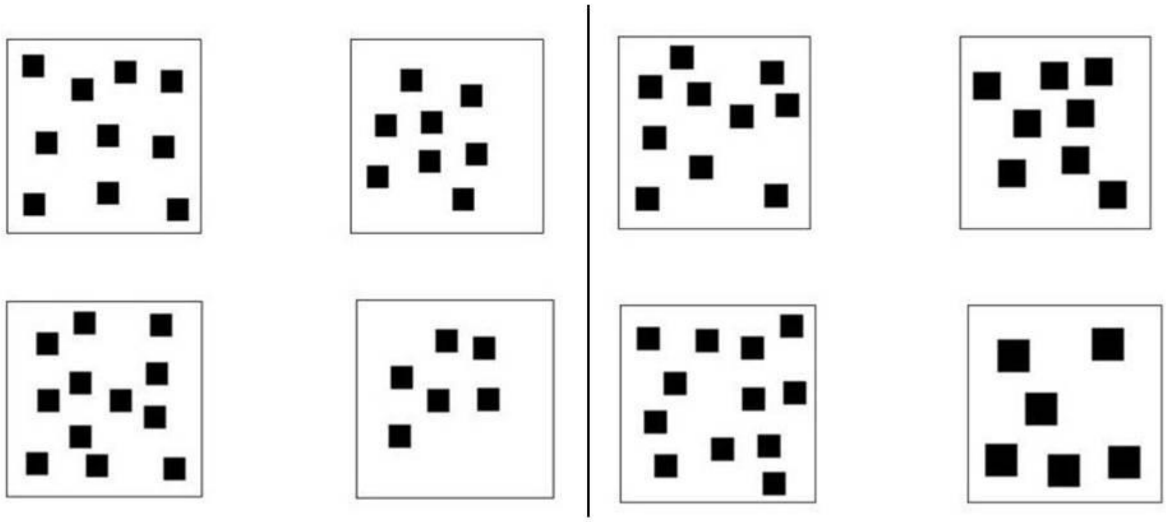


Figure 2.2. Example of the nonsymbolic comparison stimuli designed to investigate the distance effect (close and far comparison distances) with same size stimuli on the left and stimuli controlled for total surface area on the right.

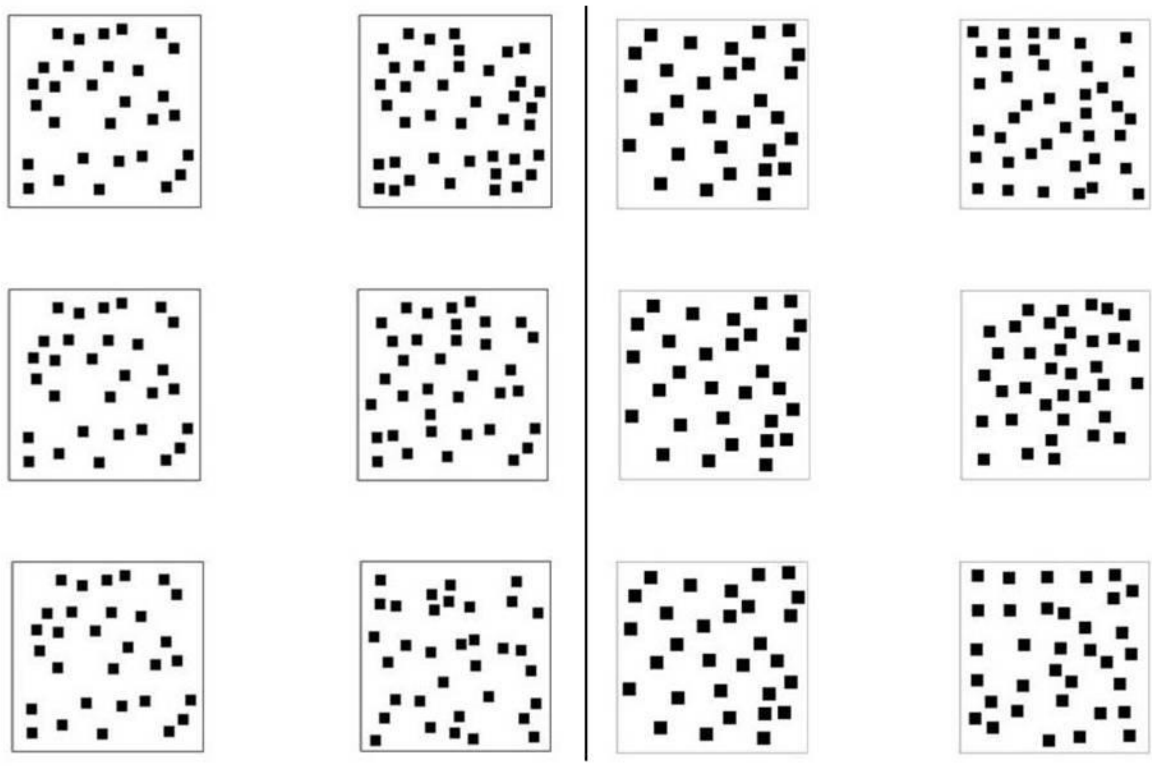


Figure 2.3. Example of the nonsymbolic comparison stimuli designed to investigate the effect of ratio (3:4, 5:6, and 7:8 comparison ratios) with same size stimuli on the left and stimuli controlled for total surface area on the right.

2.2.2.6.3.2. Effect of ratio.

This task was constructed to investigate the effect of the ratio difference between the two numerosities being compared. It included larger numerosities that ranged from 20 to 40. The ratios between the numerosities were 3:4, 5:6, and 7:8 and each of these tasks were presented separately. These three ratios were chosen after piloting six different ratios (2:3, 3:4, 4:5, 5:6, 6:7, and 7:8) as they showed the clearest results and covered a range of difficulty. The numbers 20 through to 30 were used as a baseline and the nearest whole number at each ratio was calculated. For example 20 was compared to 27 (3:4), 24 (5:6), and 23 (7:8). Figure 2.3 illustrates the task.

Table 2.3 gives a summary of the comparison tasks presented to the children and includes the short hand labels used to refer to the individual tasks in the results section.

Table 2.3.
Summary of comparison tasks presented

Comparison task	Versions	Label	Item range
Symbolic digit	Far	Digit far	1 – 9
	Close	Digit close	
Letter	Far	Latter far	a – i
	Close	Letter close	
Nonsymbolic:	Far: Same size	NS far SS	5 - 13
Effect of distance	Far: Surface area matched	NS far SA	
	Close: Same size	NS close SS	
	Close: Surface area matched	NS close SA	
Nonsymbolic:	3:4: Same size	NS 3:4 SS	20 - 40
Effect of Ratio	3:4: Surface area matched	NS 3:4 SA	
	5:6: Same size	NS 5:6 SS	
	5:6: Surface area matched	NS 5:6 SA	
	7:8: Same size	NS 7:8 SS	
	7:8: Surface area matched	NS 7:8 SA	

2.2.3. Procedure

Children were tested between April and July 2010. The children were assessed in whole class groups (i.e. 19 to 30 children). The measures were presented in three testing sessions lasting on average an hour each. The sessions were typically less than two weeks apart. The order of presentation was, where possible, kept constant: comparison tasks part 1, Raven's Matrices, one minute addition, number writing, number identification, BPVS part 1, WIAT: Numerical Operations, comparison tasks part 2, BPVS part 2. One minute subtraction was presented to the children when class time permitted. Due to time constraints in some cases the presentation order was changed (see Appendix 8 for the task presentation order of each school). Each child had their own booklet to mark their answers in and where necessary were presented with examples using the classroom whiteboard.

2.3. Results

The descriptive statistics are presented first followed by the exploration of the relationships between the measures. As children were tested as a whole class over three sessions not all children were present for all measures. The number of children who completed each task is shown for each measure.

2.3.1. Descriptive Statistics

Children's performance on the general ability measures, the three arithmetic tasks, the two specific number skill tasks and the magnitude comparison measures are presented in Table 2.4. Performance on each test is presented as a raw score and these represent the number of items correct. On the magnitude comparison tasks scores represent the number of items correct within the time limit of 30 seconds. There was a wide range of scores on all measures.

There were ceiling effects on the number knowledge tasks which required further investigation. From the histograms the number writing task showed negative skew, while the number identification task showed a more normal distribution. The number writing data were then examined by item and it was found that children were quite accurate at writing two digit numbers (percentage of children transcribing each item correctly: 12 = 89%, 19 = 79%, 20 = 82%, 37 = 69%, 63 = 67%, 100 = 81%, 152 = 23%). In fact 52% of children were able to write all five of the two digit numbers correctly (5 out of 7 items correct). Due to the low variance and ceiling effects in this measure it was decided to exclude it from further analysis.

Table 2.4.
Descriptive statistics for the group administered measures

	N	Max achievable score	Mean	SD	Min	Max
Raven's Matrices	167	34	12.25	4.13	5	23
BPVS	156	36	30.86	3.27	18	36 (3)
Numerical Operations	166	15	8.86	2.47	2	14
One minute addition	165	30	7.80	4.28	0 (4)	23
One minute subtraction[a]	79	30	6.86	3.82	0 (6)	15
Number writing	159	7	4.91	1.84	0 (3)	7 (29)
Number ID	156	8	5.08	1.59	0 (1)	8 (22)
Digit close	159	36	17.28	5.68	2	34
Digit far	164	36	20.54	6.02	5	36 (2)
Letter close	160	36	5.16	3.38	0 (10)	18
Letter far	161	36	8.59	4.69	0 (4)	24
NS close SS	163	36	12.56	4.71	0 (2)	25
NS far SS	160	36	21.91	6.68	1	36 (4)
NS close SA	161	36	11.53	4.65	1	25
NS far SA	164	36	15.12	5.79	0 (3)	36 (1)
NS 3:4 SS	164	36	13.02	5.40	0 (1)	24
NS 5:6 SS	164	36	8.29	3.90	0 (3)	20
NS 7:8 SS	161	36	11.60	4.63	0 (4)	22
NS 3:4 SA	160	36	13.17	5.02	0 (2)	28
NS 5:6 SA	159	36	11.14	4.38	1	26
NS 7:8 SA	164	36	8.73	3.67	0 (1)	20

Note. NS = nonsymbolic; close = distance of 1 or 2, far = distance of 5, 6, and 7; SS = same size stimuli, SA = surface area controlled stimuli. Figures in parentheses represent the number of children scoring at floor and ceiling.
[a] this measure was not given to all classes due to time restrictions.

2.3.2. Statistical Analyses of Comparison Tasks

2.3.2.1. Symbol comparison.

From Figure 2.4 it can be seen that on both symbolic comparison measures children gained higher scores when the digits and letters were far apart than when they were close together. Children also completed more items on the digit comparison tasks than on the letter comparison tasks. To investigate the effect of distance and symbol on children's symbolic comparison ability a two-way repeated-measures ANOVA was performed (distance: far and close, symbol: digit and letter). There was a significant main effect of distance, $F(1, 145) = 990.33$, $p < .001$, $\eta_p^2 = .87$ and a significant effect of symbol type, $F(1, 145) = 140.81$, $p < .001$, $\eta_p^2 = .49$ on children's comparison performance. There was no significant interaction suggesting that the manipulation of distance affected both symbols in the same way.

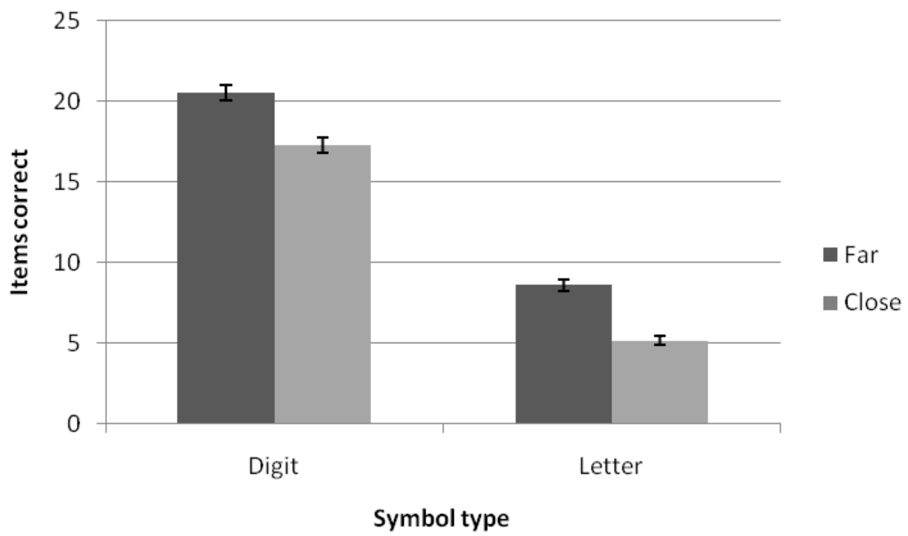


Figure 2.4. Children's performance on the symbolic comparison tasks by type of symbol and distance.

2.3.2.2. Nonsymbolic comparison: Effect of distance.

From Figure 2.5 it can be seen that children gained higher scores on both of the far comparison tasks than the close tasks. To investigate this and whether there was an effect of the type of presentation of the stimuli a two-way repeated-measures ANOVA of distance (close versus far) and presentation of the stimuli (same size versus surface area controlled) was carried out. Again there was a significant effect of distance on children's comparison performance, $F(1, 149) = 435.12$, $p < .001$, $\eta_p^2 = .75$, with children completing more correct comparisons when the distance between the numerosities was far than when it was close. There was a significant effect of the presentation of the stimuli on children's comparison performance, $F(1, 149) = 180.40$, $p < .001$, $\eta_p^2 = .55$, with children overall performing better when comparing stimuli that were the same size than when the total surface area of the stimuli was matched. There was also a significant interaction between the distance between the numerosities and the presentation of the stimuli, $F(1, 149) = 96.60$, $p < .001$, $\eta_p^2 = .39$.

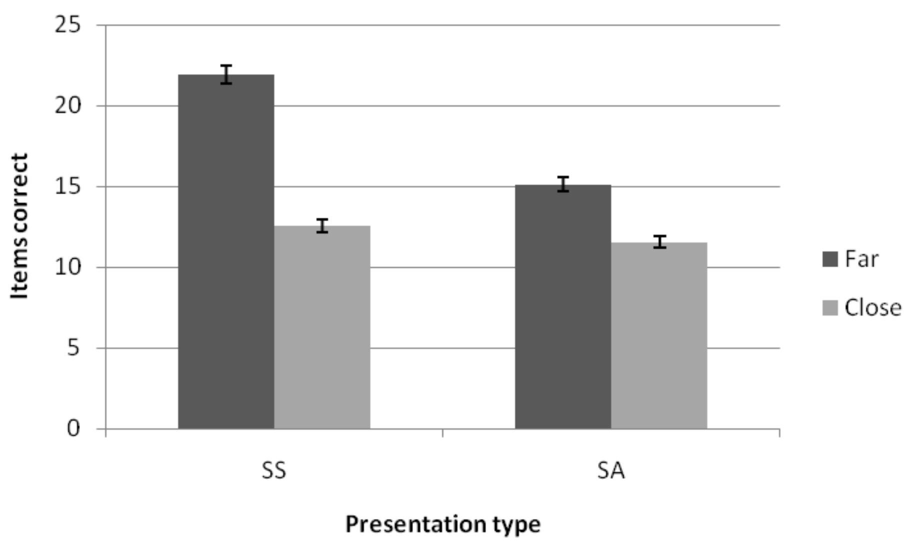


Figure 2.5. Children's performance on the nonsymbolic comparison tasks by presentation of the stimuli and distance.

To investigate the interaction effect Tukey's HSD posthoc comparisons were carried out (HSD = 1.24, $p < .05$, therefore a difference of 1.24 between two means reflects a significant difference). Four comparisons were made in total and revealed that the only comparison that was not significant was between the close SS and close SA tasks. All other means were significantly different to each other confirming that children gained higher scores when comparing items that were far apart than when they were close together regardless of presentation of the stimuli. A larger effect size was also found for the effect of distance on children's performance on the same size presented task (Cohen's d = 1.67) than the surface area presented task (d = 0.58) meaning that the distance effect is stronger for the same size stimuli.

2.3.2.3. Nonsymbolic comparison: Effect of ratio.

Figure 2.6 shows that when the stimuli presented had the same size squares children on average gained higher scores on the 3:4 ratio followed by the 7:8 ratio, whilst the lowest performance was observed at the 5:6 ratio. When the stimuli presented were matched for surface area the expected pattern was observed with children's performance decreasing as the ratio decreased (3:4>5:6>7:8). A factorial 2 (stimuli presentation: SS, SA) X 3 (ratio: 3:4, 5:6, 7:8) repeated-measures ANOVA was performed to investigate this. Mauchly's test indicated that the assumption of sphericity had been violated for the interaction term, $\chi^2(2) = 16.26$, $p < .001$. Therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .91$). On this ratio manipulated task there was no effect of presentation type of the stimuli on children's comparison performance, $F(1, 149) = 0.19$, $p = .666$, $\eta_p^2 = .00$. There was a significant main effect of ratio on children's comparison performance, $F(2, 298) = 140.04$, $p < .001$, $\eta_p^2 = .48$. There was a significant interaction effect between ratio and the presentation of the stimuli, $F(1.81, 269.92) = 80.10$, $p <.001$, $\eta_p^2 = .35$ (see Figure 2.6).

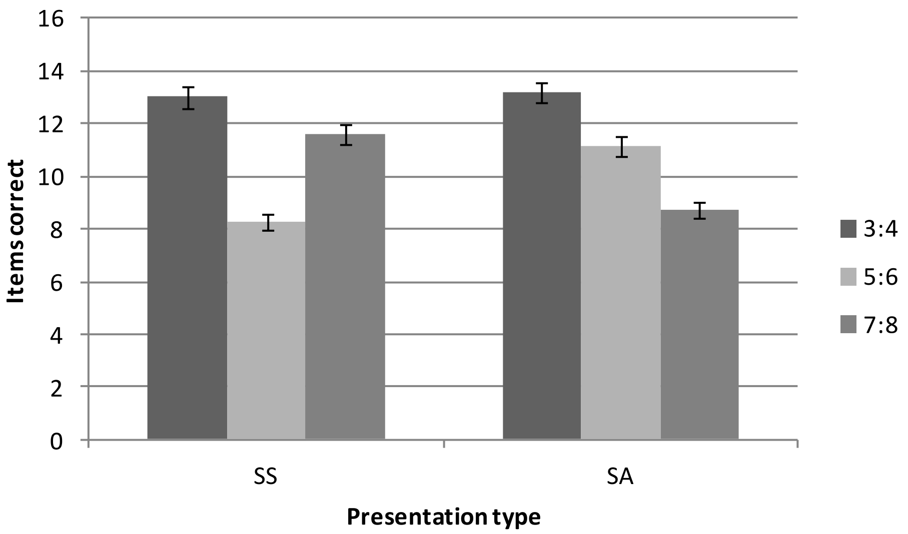


Figure 2.6. The Interaction between ratio and type of stimuli presented (where SS represents same size stimuli and SA represents surface area matched stimuli). Error bars represent standard error.

To investigate the interaction effect Tukey's HSD posthoc comparisons were carried out (HSD = 1.02, $p < .05$, therefore a difference of 1.02 between two means reflects a significant difference). Here we are interested in the comparisons within presentation type but across ratio (i.e. the ratio effect, 4 comparisons), and within ratio but across presentation type (3 comparisons) (for mean scores for each task seeTable 2.4 and Figure 2.6). All comparisons were significant except when comparing children's performance on the two 3:4 ratio tasks. Regarding the effect of ratio within stimulus presentation type the trend observed in the surface area matched condition where mean scores increased with each increase in ratio (3:4 > 5:6 > 7:8) was supported. In the same size condition as expected children gained significantly higher scores when the ratio between the numerosities was 3:4 than both 5:6 and 7:8. However, children actually gained significantly higher scores when comparing numerosities at the 7:8 ratio than the 5:6 ratio. This contradictory result could be due to the order the comparison tasks were presented to the children in the booklets, the 5:6 same size comparison task was the first large numerosity comparison task to be presented to the children and was also early in the first booklet.

2.3.3. Concurrent Relationships

To investigate the relationships between the observed possible predictor measures and arithmetic, correlations between these measures were conducted.

2.3.3.1. Correlations between measures.

Simple and partial correlations controlling for age were conducted (Table 2.5). In general the strength of the relationships were similar whether controlling or not for age. The arithmetic measures correlate moderately to strongly with each other. The two general ability measures have a very weak relationship with each other. Raven's matrices (nonverbal ability) is significantly albeit weakly associated with the three arithmetic measures, whereas vocabulary (BPVS) is not significantly correlated with one minute addition and again shows only a weak relationship with one minute subtraction and Numerical Operations. Number identification was moderately associated with all of the arithmetic measures.

All of the correlations between the magnitude comparison tasks were significant and positive suggesting that children who did better on one measure also performed better on the others. The strength of the associations ranged from weak to strong. There were low to high correlations between the measures used to investigate the distance effect. Although the strength of the relationships between the letter close task and the other measures were lower on average than the others. There were moderate correlations between the two distances (close and far) within each of the digit, letter and nonsymbolic versions of the task. The relationships between the ratio manipulated measures were moderate to high between all the tasks.

All of the magnitude comparison tasks were significantly related to performance on the Numerical Operations and one minute addition arithmetic measures. The correlations ranged from weak to moderate in strength with a trend for stronger associations with the pure speeded measure. In general the strength of the correlations between the magnitude comparison tasks and one minute subtraction were weaker and not all reached significance (though a smaller number of children completed the subtraction task).

Table 2.5.

Correlations between all group measures

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Age	.15	.27**	.15*	.29**	.27*	.17*	.22**	.26**	.22**	.05	.15	.26**	.17*	-.03	.20*	.22**	.15
2. Ravens		.17*	.33*	.35**	.38**	.29**	.29**	.40**	.41**	.23**	.32**	.29**	.24**	.25**	.24**	.26**	.12
3. BPVS	.14		.36**	.20*	.31**	.43**	.21*	.18*	.18*	.14	.21*	.27**	.23**	.16	.20*	.11	.17*
4. WIAT: Numerical Operations	.31**	.33**		.51**	.66**	.47**	.39**	.44**	.30**	.26**	.41**	.37**	.39**	.33**	.36**	.28**	.25**
5. One minute addition	.33**	.14	.49**		.65**	.40**	.43**	.52**	.38**	.42**	.38**	.47**	.40**	.34**	.39**	.32**	.34**
6. One minute subtraction[a]	.36**	.25*	.66**	.62**		.48**	.45**	.42**	.30**	.28*	.31**	.16	.38**	.08	.40**	.33**	.25*
7. Number identification	.27**	.40**	.45**	.37**	.46**		.19*	.31**	.26**	.25**	.32**	.35**	.23**	.25**	.30**	.22**	.24**
8. Digit far	.27**	.16	.37**	.40**	.42**	.16*		.59**	.41**	.28**	.55**	.49**	.56**	.45**	.58**	.45**	.56**
9. Digit close	.38**	.12	.42**	.48**	.38**	.28**	.57**		.54**	.32**	.69**	.58**	.50**	.64**	.71**	.66**	.51**
10. Letter far	.39**	.13	.28**	.34**	.25*	.24**	.38**	.51**		.49**	.47**	.42**	.37**	.55**	.37**	.39**	.37**
11. Letter close	.22**	.13	.26**	.42**	.27*	.24**	.28**	.32**	.50**		.34**	.35**	.38**	.34**	.30**	.18*	.29**
12. NS far SS	.30**	.18*	.39**	.35**	.28*	.31**	.54**	.68**	.45**	.33**		.55**	.49**	.68**	.58**	.62**	.48**
13. NS close SS	.27**	.22**	.35**	.43**	.09	.32**	.46**	.55**	.38**	.35**	.54**		.50**	.56**	.55**	.47**	.57**
14. NS far SA	.22**	.19*	.38**	.37**	.35**	.21*	.54**	.48**	.34**	.38**	.48**	.48**		.46**	.46**	.33**	.51**
15. NS close SA	.26**	.17*	.33**	.36**	.10	.26**	.47**	.68**	.57**	.34**	.70**	.59**	.48**		.49**	.53**	.50**
16. NS 3:4 SA	.21*	.16*	.34**	.35**	.36**	.27**	.56**	.69**	.34**	.30**	.57**	.52**	.44**	.50**		.68**	.50**
17. NS 5:6 SA	.23**	.06	.26**	.28**	.29*	.19*	.43**	.64**	.36**	.17*	.61**	.43**	.31**	.55**	.67**		.43**
18. NS 7:8 SA	.10	.13	.23**	.31**	.22	.22**	.54**	.49**	.35**	.28**	.47**	.55**	.49**	.51**	.49**	.41**	

Note. [a] this measure was not presented to all children due to time restrictions.

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

2.3.3.2. Structural Equation Modeling (SEM).

The main aims of this study were to investigate the underlying constructs of the comparison tasks and to assess whether children's performance on these tasks was related to their arithmetic achievement. To address the first aim and investigate the possible latent factors of the comparison tasks an ideal method of analysis is confirmatory factor analysis (CFA). With the large sample size of this study Structural Equation Modeling (SEM) techniques could be employed. This method enables the investigation of how the comparison tasks are related to each other by forming latent factors that represent underlying constructs, and also tests how these proposed models fit with the data. This is an advantage of using SEM rather than using the strength of correlations to construct composite scores. The relationships between these latent factors and the other measures (arithmetic achievement, general ability, number identification) can be assessed. The CFA can also be used as the basis of a path model to predict individual differences in children's arithmetic achievement.

2.3.3.2.1. SEM methods.

All SEM analyses were conducted using M*plus* (Muthén & Muthén, 1998-2011) and missing values were handled with Full Information Maximum Likelihood estimation. Before presenting the results an interpretation of the models is provided. Measures that are presented in rectangular boxes are observed variables, while those presented in an ellipse reflect latent variables/factors (these are the underlying constructs that are believed to result in the performance on the observed variables). Single-headed arrows from the latent to the observed variables, or in a path model from the CFA to the outcome variables (latent or observed) reflect a causal path and the values connected to these are the factor loadings. The values presented are the standardised regression coefficients; these are the same as beta weights from a regression analysis. Therefore a change of one standard deviation in the predictor variable (X) results in a difference of the beta weight (e.g. β = .52) on the outcome (Y). As the values reported are standardised the contribution of the predictors can be compared. Two-headed arrows between the variables reflect true-score correlations between constructs. On the final path model figures the single headed arrows pointing at the outcome latent variable represent the residual and therefore the unexplained variance in the measure. A significant path is represented by a solid line, whereas a non significant path is represented by a dashed line.

To assess the goodness of fit between the proposed model and the sample data the program reports multiple statistics, four will be presented here. The first value is the result of a chi-square difference test; this represents the difference between the covariance matrix of the sample data and the covariance matrix implied by the model (Byrne, 2012). The disadvantage of using chi-square as an indication of model fit is that it is sensitive to sample size, and with SEM being a large sample technique then it will be more likely that the chi-square difference test is significant (ideally we would want a non-significant result as this would indicate that there is no difference between the covariance matrix of the model and from the sample data). Therefore other fit indices are used to assess the hypothesized model fit. With regards to the Root Mean Square Error of Approximation (RMSEA), which compensates for the effect of model complexity, it has been suggested that a value of less than .06 indicates a good fit (Hu & Bentler, 1999), whereas values of less than .08 have been considered an acceptable fit, and values of less than .05 indicating a good fit (Browne & Cudeck, 1992). The Comparative Fit Index (CFI) value ranges from 0 to 1.00, the closer the value to one the better the model fit. Hu and Bentler (1999) recommend that the value of a well-fitting model is greater than .95. The final reported model fit indices is the Standardized Root Mean Residuals (SRMR), ideally we are looking for a value of less than .05 (Byrne, 2012).

2.3.3.2.2. CFA of comparison measures.

Method: First a confirmatory factor analysis (CFA) was conducted to assess the relationships between the comparison tasks. This was to investigate whether the three types of comparison task (digit, letter and nonsymbolic) that were used in this study would represent one single factor (general comparison ability), two factors (magnitude comparison and letter comparison or symbolic and nonsymbolic comparison), or even three separate factors (digit, nonsymbolic and letter).

Children completed multiple comparison tasks, however the presentation order of the tasks may have affected children's performance on the nonsymbolic ratio manipulated tasks. The typical ratio effect was found on the stimuli presented that had the same total surface area (3:4 > 5:6 > 7:8), however the ratio effect observed when the comparison stimuli were the same size was not in the expected direction (3:4 > 7:8 > 5:6). Therefore these three comparison tasks with the same size manipulation were not included in further analysis. To simplify the model further and reduce the number of free parameters to be estimated only two of the smaller numerosity nonsymbolic tasks were included in the analysis. As the surface area matched stimuli from the ratio manipulation will be used, the same size stimuli were chosen from the distance effect comparison tasks as to include both types of stimuli manipulation. All of the four symbolic comparison tasks were entered into the analysis. Therefore the following measures were entered into the CFA: digit far, digit close, letter far, letter close, NS far SS, NS close SS, NS 3:4 SA, NS 5:6 SA, and NS 7:8 SA.

Results: The first analysis was run to investigate whether all comparison tasks (digit, nonsymbolic and letter) represent one latent factor. It is possible that all three types of comparison task tap a general comparison construct which is reflected in the ability to compare two items and choose one of them. Figure 2.7 shows the one factor CFA, this model did not provide an adequate fit to the data, χ^2 (27) = 73.644, p = <.001, *RMSEA* = .100 (90% *CI* = .073-.128), *CFI* = .933, *SRMR* = .055.

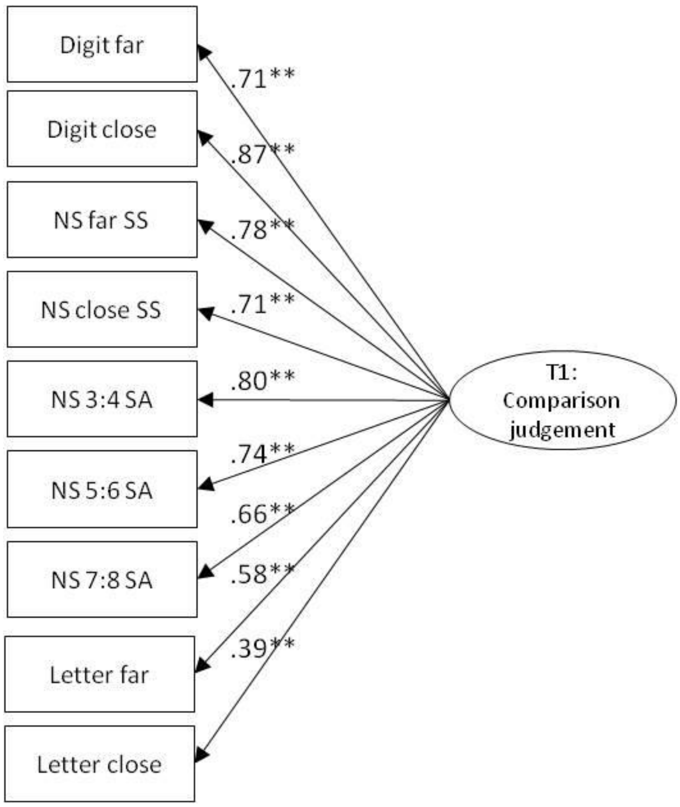


Figure 2.7. One factor CFA of comparison tasks (Time 1). Model fit: χ^2 (27) = 73.644, p = <.001, *RMSEA* = .100 (90% *CI* = .073-.128), *CFI* = .933, *SRMR* = .055.
* p < .05, ** p < .01.

It was suggested that the comparison tasks could form two latent factors, either a magnitude (digit and nonsymbolic) comparison factor and a separate letter comparison factor, or another possible latent two factor structure could emerge with a symbolic comparison construct (both digit and letter) and a separate nonsymbolic factor. The first two factor solution was tested and is presented in Figure 2.8, the model provided an acceptable fit to the data, χ^2 (26) = 53.567, *p* = .001, *RMSEA* = .078 (90% *CI* = .048-.108), *CFI* = .960, *SRMR* = .041. The second two factor solution of symbolic (digit and letter) and nonsymbolic latent factors did not provide an acceptable fit to the data, χ^2 (26) = 73.241, *p* < .001, *RMSEA* = .102 (90% *CI* = .075-.131), *CFI* = .932, *SRMR* = .054 (see Figure 2.9). The relationship between the two constructs was also very high.

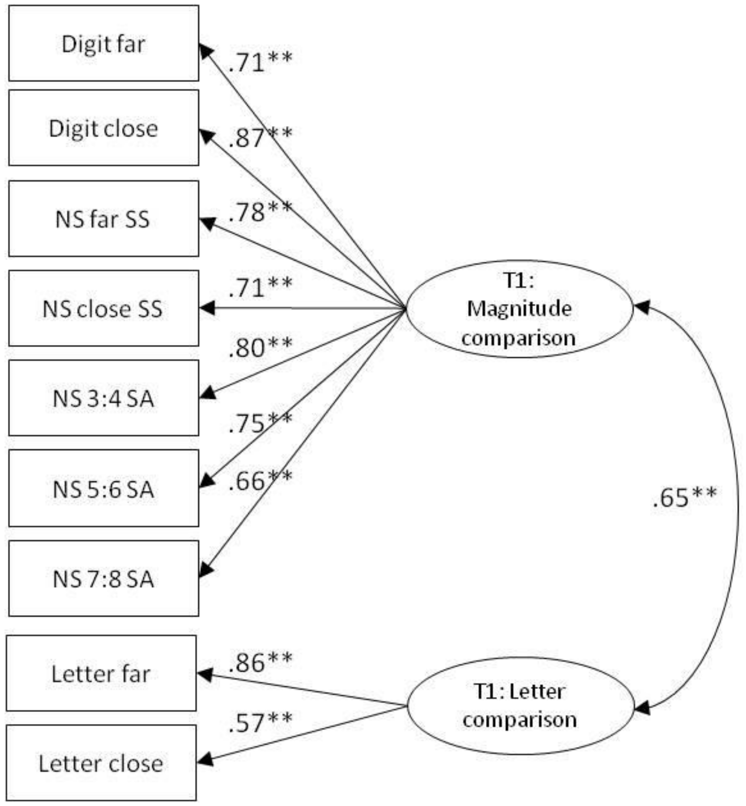


Figure 2.8. Two factor CFA of comparison tasks: Magnitude comparison and letter comparison (Time 1). Model fit: χ^2 (26) = 53.567, *p* = .001, *RMSEA* = .078 (90% *CI* = .048-.108), *CFI* = .960, *SRMR* = .041.
* *p* < .05, ** *p* < .01.

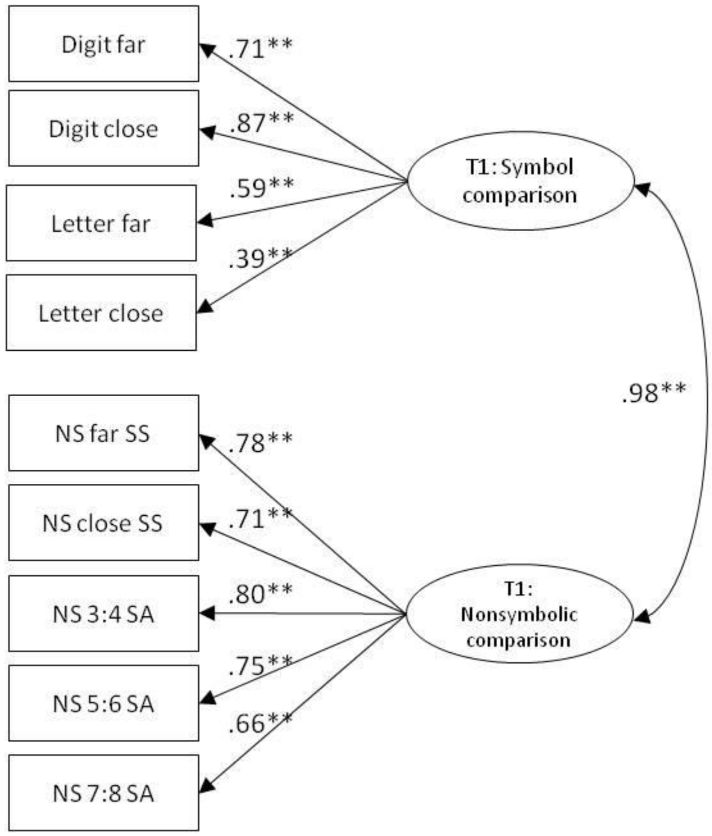


Figure 2.9. Two factor CFA of comparison tasks: Symbol comparison and nonsymbolic comparison (Time 1). Model fit: χ^2 (26) = 73.241, $p < .001$, *RMSEA* = .102 (90% *CI* = .075-.131), *CFI* = .932, *SRMR* = .054.
* $p < .05$, ** $p < .01$.

The final possible structure of the comparison tasks could be a three factor solution. Each comparison task could form its own latent variable due to the different items being compared (digit, letter and nonsymbolic) and may rely on a different mechanism and way to complete the task. This solution (see Figure 2.10) provided a reasonable fit to the data, χ^2 (24) = 51.595, p = <.001, *RMSEA* = .082 (90% *CI* = .051-.112), *CFI* = .960, *SRMR* = .042 however the coefficient between the digit and nonsymbolic latent variables was $r = 1.01$, $p < .001$. On the technical side this causes an issue for the analysis, whereas on the theoretical side suggests that these two constructs are in fact the measuring the same underlying skill.

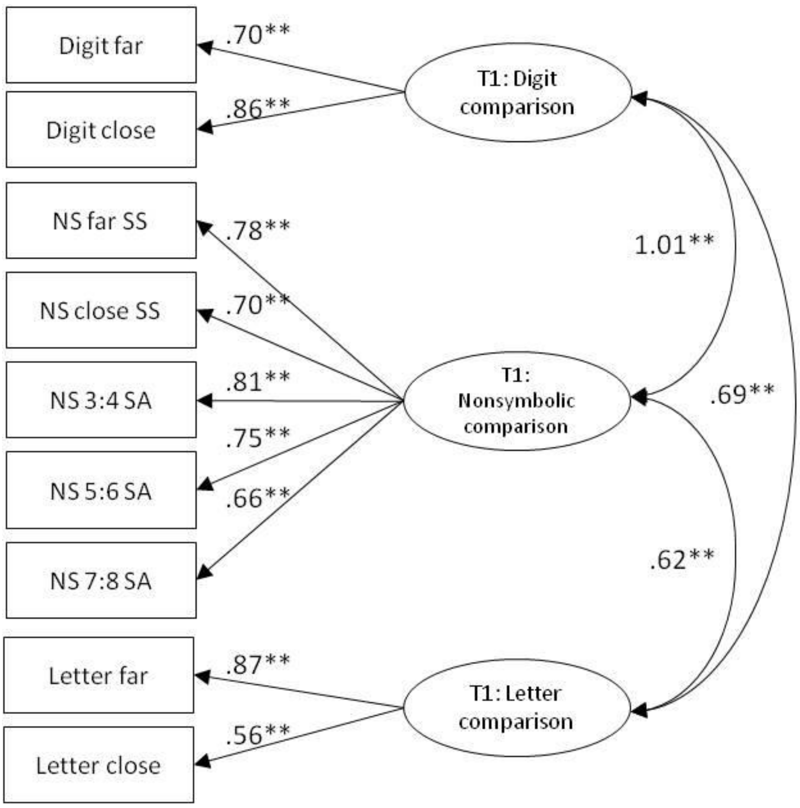


Figure 2.10. Three factor CFA of comparison tasks (Time 1). Model fit: χ^2 (24) = 51.595, *p* = <.001, *RMSEA* = .082 (90% *CI* = .051-.112), *CFI* = .960, *SRMR* = .042.
* *p* < .05, ** *p* < .01.

Referring back to the model fit indices suggested as a guideline earlier, the one factor solution and the two factor solution with symbol comparison (digit and letter) and nonsymbolic comparison factors should be discounted due to the large *RMSEA* values (both greater than .08), larger than ideal *SRMR* values (greater than .05), and a *CFI* of less than .95. This leaves the two factor solution with magnitude (digit and nonsymbolic) and letter comparison latent factors and the three factor solution where every type of comparison task forms its own factor as the best fitting model solutions. Both models fit the data equally well and a chi-square difference test revealed that they were not significantly different ($\Delta\chi^2$ = 1.97, df = 2, *p* > .05). However, in the three factor solution the relationship between the symbolic digit and nonsymbolic factors is *r* = 1.01, this would suggest that these two factors are not separate underlying constructs. This is an interesting finding that the best fitting model is one where digit and nonsymbolic comparison tasks load on the same factor (magnitude comparison) whereas letter comparison forms a separate factor. There was a significant (moderate) relationship between the two latent variables which could reflect the method variance of the measures. This CFA (Figure 2.8) will be used to represent the underlying constructs of the comparison measures and will form the basis of the next stage of analysis.

2.3.3.2.3. Concurrent relationships between comparison tasks, arithmetic (WIAT), nonverbal ability, vocabulary, number ID, and age.

The next step was to assess the relationships between the comparison tasks, arithmetic (WIAT), nonverbal ability (Raven's), vocabulary (BPVS), number ID, and age. The comparison CFA (Figure 2.8) was included alongside the other constructs measured. In this model, the WIAT arithmetic test was divided into three indicators (by taking every third item on the test) to allow for the estimation of a latent arithmetic variable. Age was included in the analysis to control for any maturational differences between the children, along with nonverbal ability and vocabulary as any relationships between comparison ability and arithmetic achievement could be related to common variance due to individual differences in general ability. Because nonverbal ability, vocabulary and number identification were each assessed by only one indicator (Raven's, BPVS and Number ID), these indicators were pre-specified with an error reflecting the reliability of these variables calculated on the sample (to avoid distortions caused by measurement error). This model, shown in Figure 2.11, provided a good fit to the data, $\chi^2(87) = 121.834$, $p = .008$, *RMSEA* = .048 (90% *CI* = .025-.067), *CFI* = .965, *SRMR* = .044. All of the measures were significantly related to each other at this time point. With age, nonverbal ability and vocabulary controlled there were still significant relationships between WIAT arithmetic and the two comparison factors (magnitude and letter comparison), which were moderate in strength. There was also a significant association between arithmetic and number identification.

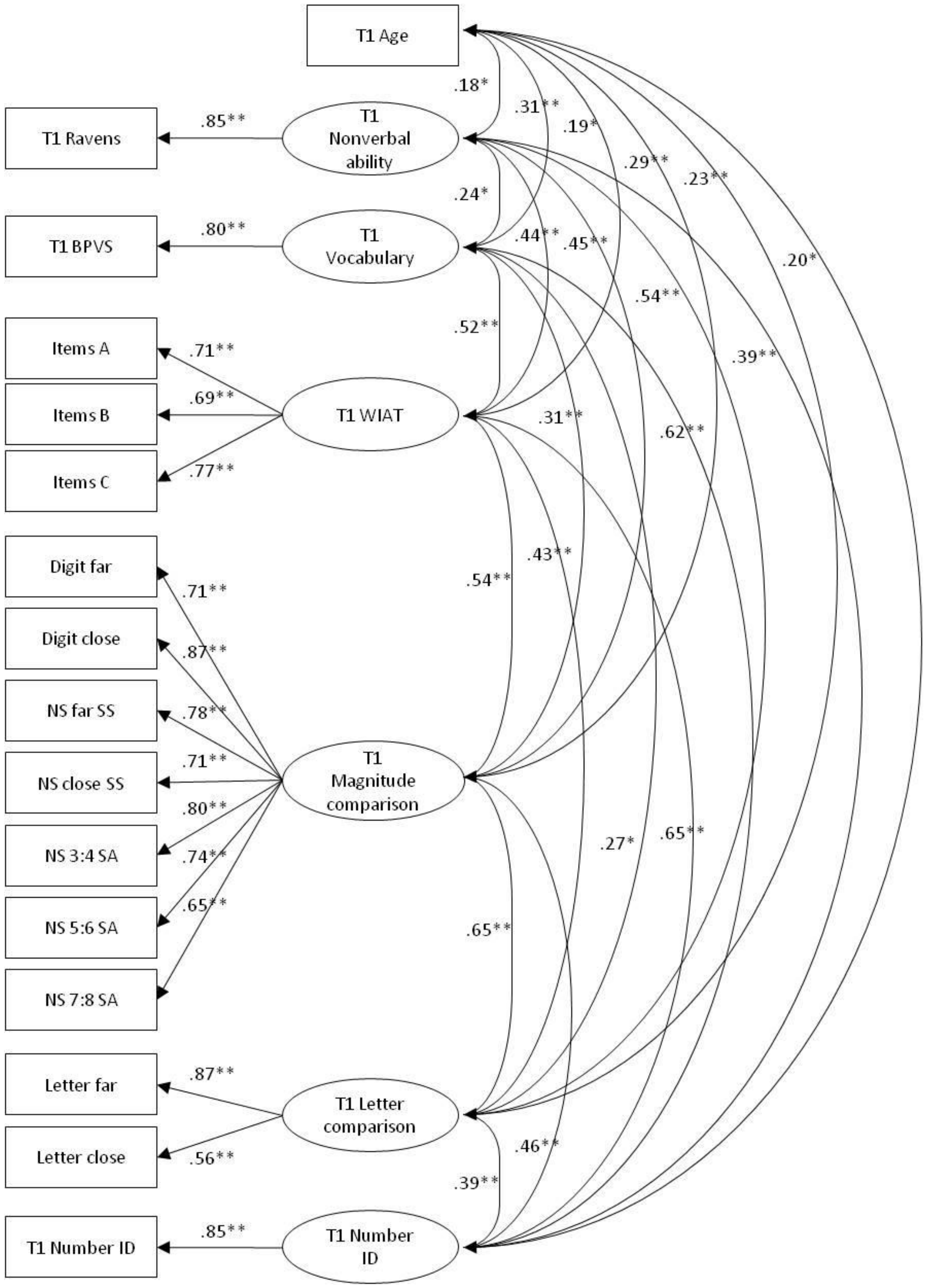


Figure 2.11. CFA including WIAT Numerical Operations, comparison, age, nonverbal ability, vocabulary and number identification (Time 1). Model fit: χ^2 (87) = 121.834, *p* = .008, *RMSEA* = .048 (90% *CI* = .025-.067), *CFI* = .965, *SRMR* = .044.
* *p* < .05, ** *p* < .01.

2.3.3.2.4. Predicting WIAT arithmetic achievement from comparison ability, nonverbal ability, vocabulary, number ID, and age (concurrently).

A CFA allows for the exploration of the relationships between the measures whilst controlling for the other variables but it does not allow for the investigation of which measures predict individual differences in children's arithmetic achievement. To investigate whether any of the measures predict individual differences in children's arithmetic achievement a path analysis was carried out. It should be noted that here we are investigating the concurrent relationships so we cannot be sure of the direction of the relationships. First, a new CFA was run removing Time1 Numerical Operations (as this will now be the predicted variable) to check the fit of the data and is illustrated in Figure 2.12. The model provided an acceptable fit to the data, χ^2 (54) = 85.783, *p* = .004, *RMSEA* = .058 (90% *CI* = .033-.081), *CFI* = .960, *SRMR* = .042. A path model was then run with children's performance on Numerical Operations as the outcome variable. Figure 2.13 illustrates the path model, which provided a good fit to the data, χ^2 (87) = 121.834, *p* = .008, *RMSEA* = .048 (90% *CI* = .025-.067), *CFI* = .965, SRMR = .044. The coefficients of the measurement model (the CFA) are not presented here for clarity but are providing in Appendix 9, it should be noted that the values are almost identical to those in the CFA (Figure 2.12). There were two significant predictors of variance in children's arithmetic scores; magnitude comparison and number identification (54% of variance was explained). This is an interesting finding that even with age, verbal ability and vocabulary in the model, children's ability to compare the magnitude of two items predicted individual differences in their arithmetic scores. Whilst children's magnitude comparison ability was a significant predictor of variance, their ability to compare letters was not. Age, nonverbal ability and vocabulary knowledge were not found to be significant predictors of individual differences in children's arithmetic achievement. At this age it looks like children's ability to identify symbolic numbers is a slightly stronger predictor of variance in arithmetic scores than magnitude comparison performance.

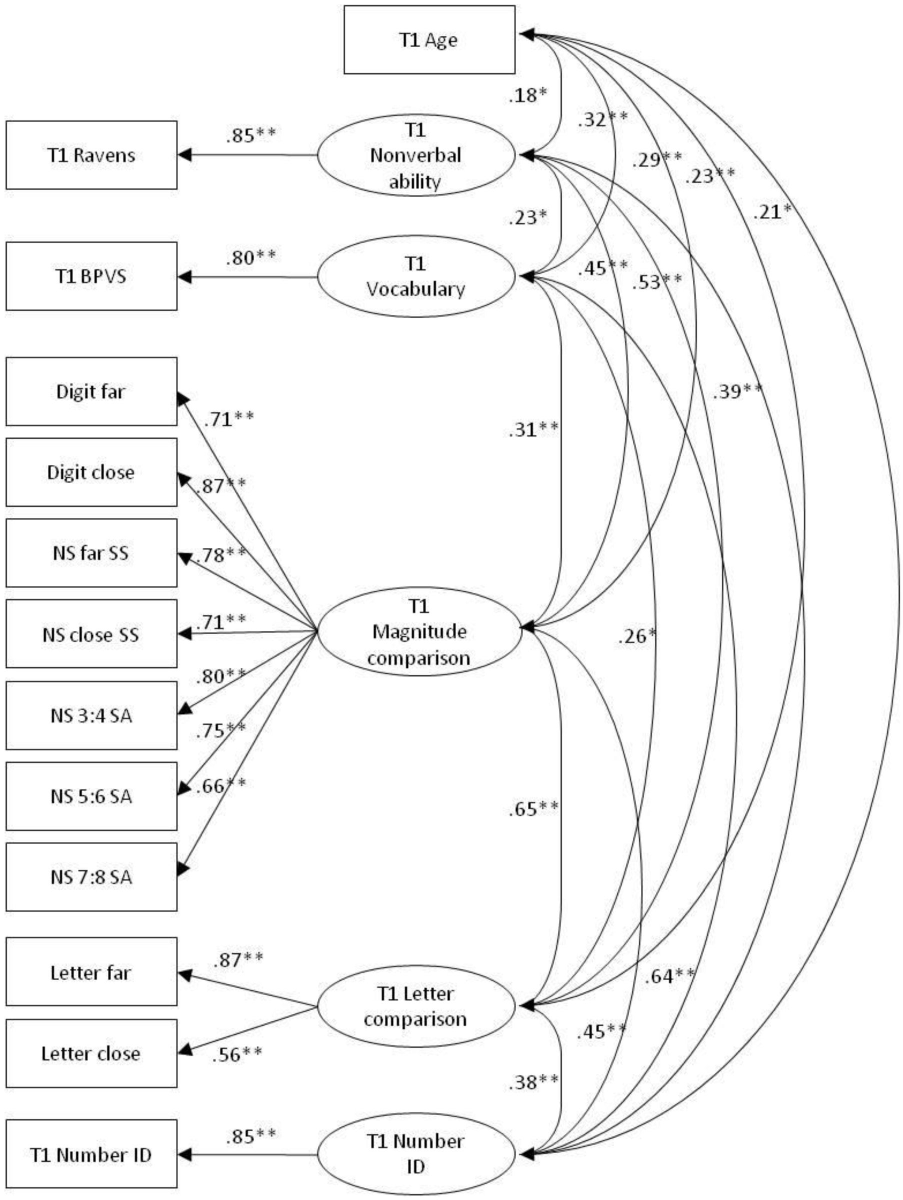


Figure 2.12. CFA of Time 1 age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification. Model fit indexes: χ^2 (54) = 85.783, *p* = .004, *RMSEA* = .058 (90% *CI* = .033-.081), *CFI* = .960, *SRMR* = .042.
* *p* < .05, ** *p* < .01.

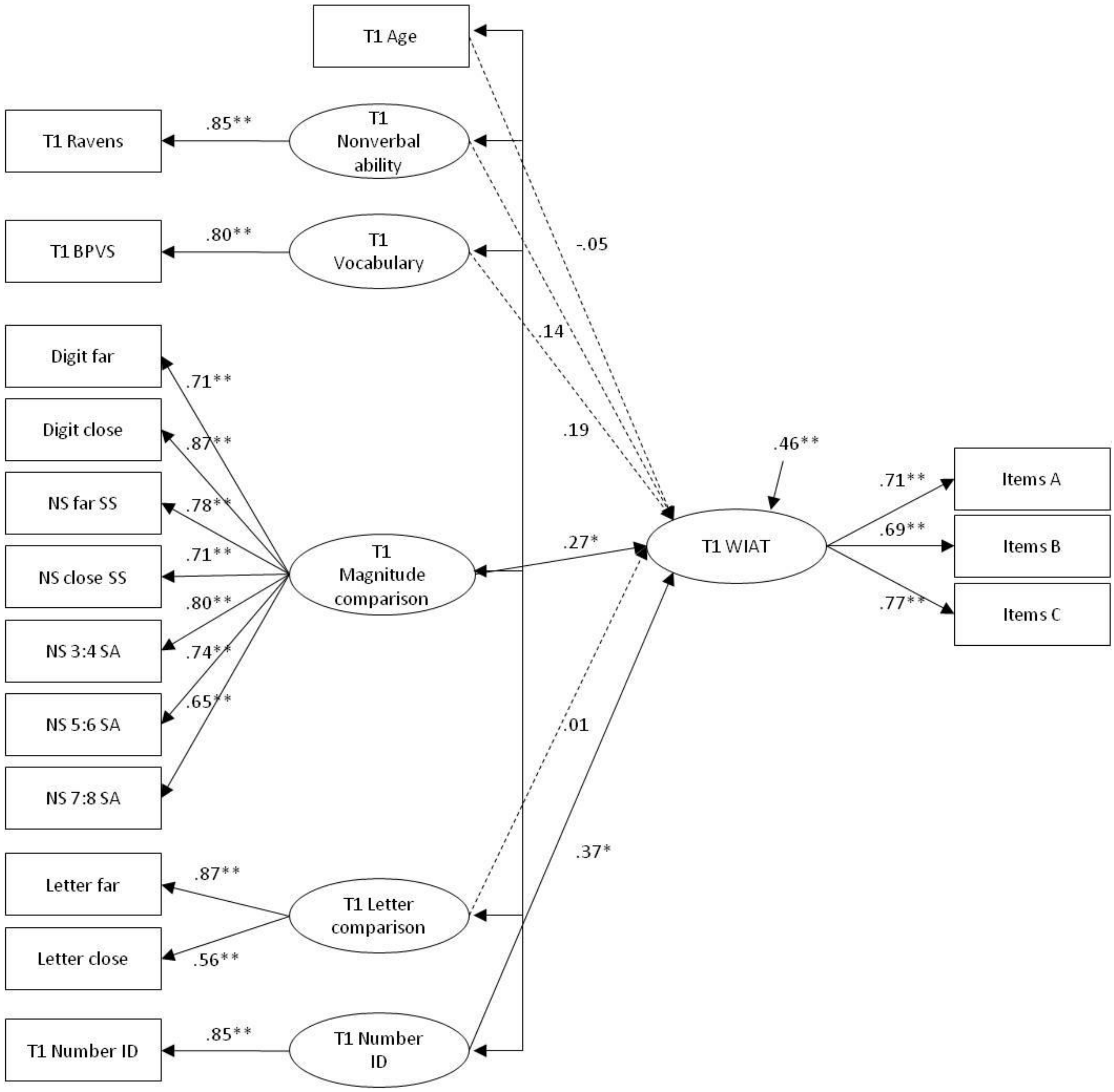


Figure 2.13. Path model predicting WIAT Numerical Operations from age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification (Time 1). Model fit: χ^2 (87) = 121.834, *p* = .008, *RMSEA* = .048 (90% *CI* = .025-.067), *CFI* = .965, SRMR = .044.

* *p* < .05, ** *p* < .01.

2.3.3.2.5. Concurrent relationships between comparison tasks, speeded arithmetic, nonverbal ability, vocabulary, number ID, and age.

To assess the whether the same findings would hold with an alternate measure used to assess arithmetic achievement, the analysis was repeated with the one minute addition test. As this was a time limited measure (one minute to complete as many addition problems as possible) children completed different numbers of items, therefore the measure was entered into the factor analysis as an observed variable (rather than creating a latent variable as with the WIAT arithmetic measure). A new CFA was run including one minute addition (T1 1 min+) alongside the comparison CFA, nonverbal ability (Ravens matrices), vocabulary (BPVS), number identification and age (see Figure 2.14). This model also provided an acceptable fit to the data, χ^2 (61) = 107.560, $p < .001$, *RMSEA* = .066 (90% *CI* = .045-.087), *CFI* = .947, *SRMR* = .043. There were significant moderate correlations between the speeded arithmetic measure (one minute addition) and both of the comparison tasks, as was seen with the untimed arithmetic measure. A significant correlation was found between number identification and speeded arithmetic but this was weaker than the relationship with WIAT arithmetic (r = .47 versus r = .65).

2.3.3.2.6. Predicting speeded arithmetic achievement from comparison ability, nonverbal ability, vocabulary, number ID, and age (concurrently).

To investigate whether any of the measures predicted individual differences in children's speeded arithmetic achievement a path analysis was carried out. The CFA depicted in Figure 2.12 (including the variables nonverbal ability, vocabulary, magnitude comparison, letter comparison, number identification and age) was again used as the measurement model, to which one minute addition was regressed on to. Figure 2.15 illustrates the path model, which provided an acceptable fit to the data, χ^2 (61) = 107.560, $p < .001$, *RMSEA* = .066 (90% *CI* = .045-.087), *CFI* = .947, *SRMR* = .043. The coefficients of the measurement model (the CFA) are not presented here for clarity but are provided in Appendix 10. There were three significant predictors of variance in children's speeded arithmetic scores; magnitude comparison, number identification, and age. In total the variables accounted for 40% of the variance in children's speeded addition scores; this is less than the total amount of variance explained in children's WIAT arithmetic scores. As with untimed arithmetic, even with age, verbal ability and vocabulary in the model, children's ability to compare the magnitude of two items (digits and nonsymbolic numerosities) predicted individual differences in their arithmetic scores. Again children's ability to identify symbolic numbers was a significant predictor of variance in their arithmetic scores. It is interesting that when exploring the predictors of variance in achievement on the speeded arithmetic task, age was a significant predictor, although this explained the least amount of variance of the three significant predictors. Children's ability to compare letters, nonverbal ability and vocabulary knowledge were not found to be significant predictors of individual differences in children's speeded arithmetic scores.

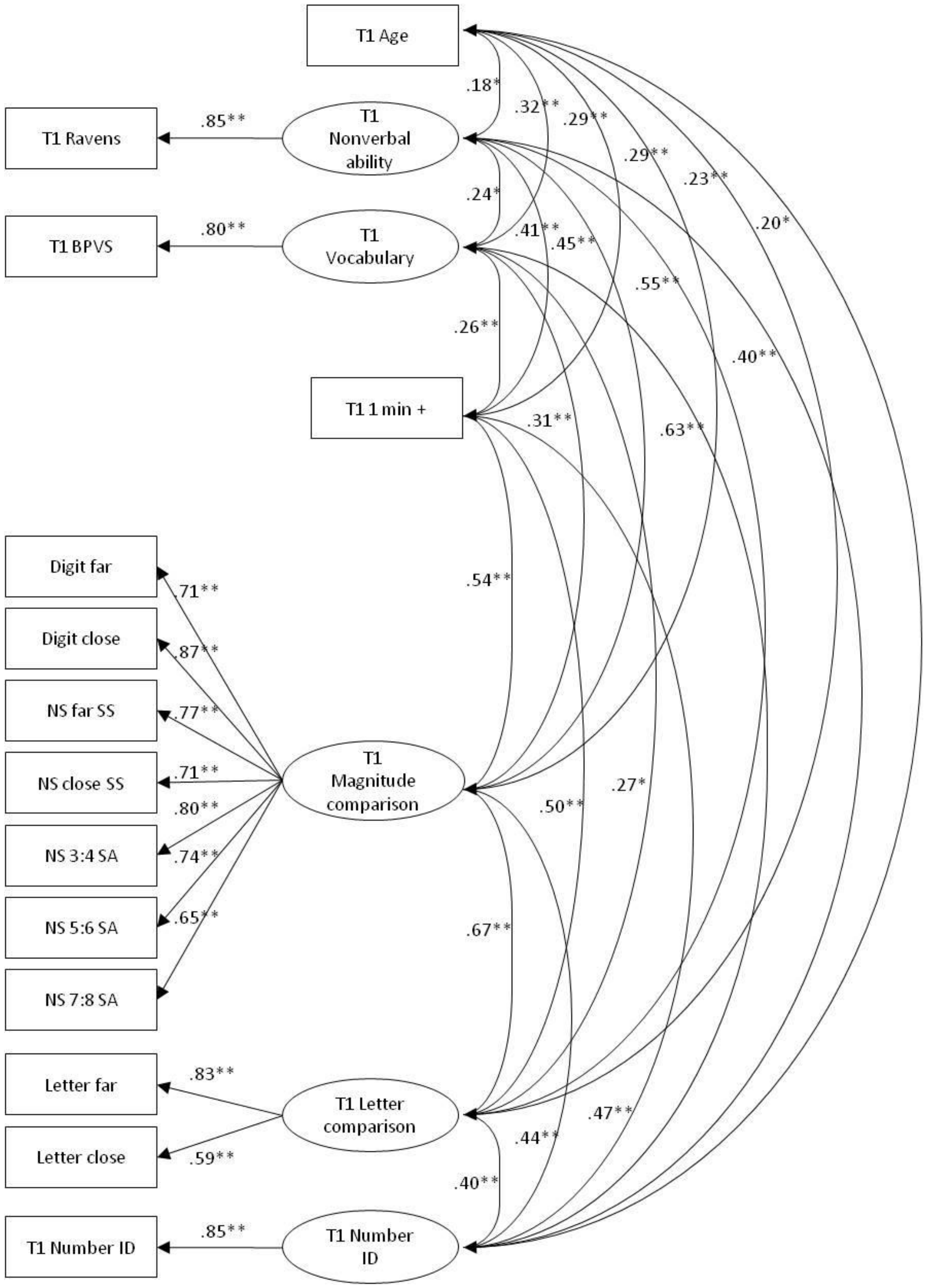


Figure 2.14. CFA including 1 minute addition, comparison, age, nonverbal ability, vocabulary and number identification (Time 1). Model fit: χ^2 (61) = 107.560, p < .001, *RMSEA* = .066 (90% *CI* = .045-.087), *CFI* = .947, *SRMR* = .043.

* p < .05, ** p < .01.

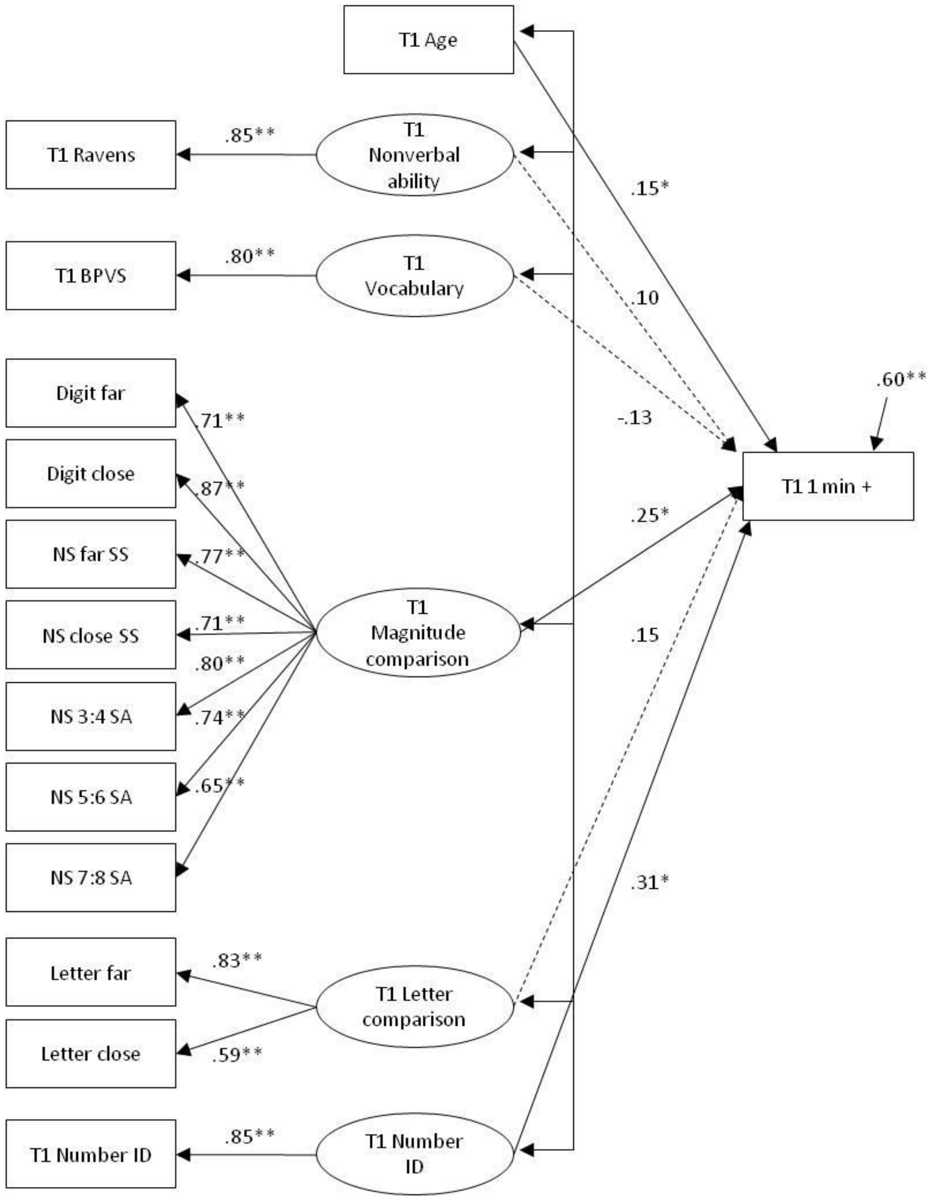


Figure 2.15. Path model predicting 1 minute addition from age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification (Time 1). Model fit: χ^2 (61) = 107.560, $p < .001$, *RMSEA* = .066 (90% *CI* = .045-.087), *CFI* = .947, *SRMR* = .043.

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

2.4. Discussion

2.4.1. Summary of Study

The main aims of the current chapter were to assess the underlying latent factors that different comparison tasks may have in common and to investigate the relationships with arithmetic achievement in children who were at an early stage of formal schooling. To answer these questions comparison tasks with a similar presentation and task demands were completed by the children and sophisticated data analysis techniques (SEM) were used to analyse the data.

2.4.2. Performance on Comparison Tasks

An effect of distance on comparison performance was observed in all of the three different categories of comparison task (digit, nonsymbolic and letter), which replicates the classic distance effect results seen in the literature (e.g. Moyer & Landauer, 1967; Holloway & Ansari, 2009; Jou & Aldridge, 1999) but in a group presentation design. This is interesting that while a distance effect was observed in all three types of comparison task, the letter and magnitude (digit and nonsymbolic) tasks loaded on different latent factors in the CFA.

An alternate way to investigate the effect of distance on comparison performance, which takes general problem size into account, is to manipulate the ratio difference between the numerosities to be compared. An effect of ratio was observed in the nonsymbolic comparison task for larger numerosities (i.e. ranging from 20 to 40). The expected pattern of results with children's scores decreasing as the ratio between the comparison items becomes more difficult (i.e. 3:4 > 5:6 > 7:8) was found when the stimuli presented were matched on the total surface area, which supports previous literature (Halberda et al., 2008). However, the pattern of results when the stimuli presented were the same size was surprising, while children gained significantly higher scores on the 3:4 ratio task than both the 5:6 and 7:8 ratio tasks, they gained the lowest scores on the 5:6 ratio (3:4 > 7:8 > 5:6). This contradictory result could be due to the order the comparison tasks were presented to the children in the booklets, the 5:6 comparison task where the stimuli were the same size was the first large numerosity comparison task to be completed by the children and was also early in the first booklet. This could have led to the poor performance on this task compared to the other ratios. Only the tasks where the typical pattern was found were used in the subsequent analysis.

The replication of the distance and ratio effects on children's comparison performance show that these novel (group presented paper and pencil) versions of the comparison tasks that are typically used in the literature are working. With the exception of the same size nonsymbolic stimuli used to investigate the ratio effect, children are processing these tasks in the same way as children/adults tested in previous studies when these measures are presented to them individually on a computer.

There was a main effect of presentation type of the stimuli on the nonsymbolic task designed to investigate the effect of distance (small numerosity), with children gaining higher scores when the stimuli were the same size than when the total surface area was matched. An interaction with distance was also observed with children gaining similar scores when the comparison items were close together but higher scores on the same size task (than when the stimuli had the same total surface area) when the comparison items were far apart. This suggests that the effect of distance was stronger when the stimuli were the same size than when they were matched on the total surface area (reflected in the larger effect size for the effect of distance when children were comparing stimuli of the same size than area matched). When the stimuli were of the same size the surface area increased proportionally with the numerosity, so children may have been using additional perceptual cues and not just numerosity to discriminate between the two (e.g. Rousselle, Palmers & Noel, 2004). This may only be true when numerosities are far apart as the perceptual difference in surface area may be more obvious than the numerical difference. When the numerosities are closer together children may indeed rely on numerical rather than perceptual cues to discriminate. If nonsymbolic comparison is found to be a significant predictor of arithmetic achievement, additional nonsymbolic comparison tasks may be required that include more stringent controls for perceptual variables (i.e. varying item size), to exclude the possibility that the relationship is with area judgement and not numerosity judgements (Chapter 6).

2.4.3. Relationship between Comparison Tasks

A main aim of the current chapter was to assess the relationships between the three different types of comparison task to investigate the possible underlying latent factors of the tasks. The best fitting model to the data was a two factor solution, with the digit and nonsymbolic comparison tasks loading on the same latent factor, and a separate letter comparison latent factor. This suggests that the digit and nonsymbolic comparison tasks are tapping the same underlying construct – magnitude comparison. This underlying factor is involved when comparing the size or magnitude of two items, be it two nonsymbolic numerosities or Arabic numerals that allows us to more simply represent numerosity. Although the letter comparison measures had a very similar presentation and task demands they loaded on a separate factor to the other comparison tasks. This suggests that the processes involved in comparing the digit and nonsymbolic items are different to those involved in comparing letters of the alphabet. The data suggest that the digit and letter comparison tasks are not tapping a suggested symbol comparison construct despite the fact that they both involve comparing learned symbols and high frequency items that are learnt by heart. In addition, both have an ordinal structure (count sequence and alphabet string). However, numbers are also cardinal and a cardinal value represents the number/quantity of items in a set. The digit (and nonsymbolic) comparison tasks therefore involve more than just comparing the order of items but also the quantities they represent.

The two comparison latent variables, magnitude and letter comparison, have a moderate-strong relationship but the model did not provide a good fit to the data when all comparison tasks were suggested to represent one single factor. This suggests that the magnitude comparison tasks are tapping something more than the ability to compare two items and choose one of them. The relationship between the latent factors could represent the shared method variance of the tasks. All comparison tasks were presented (and scored) in the same way, therefore they involved the same task demands so the relationship could reflect the speed and accuracy with which comparisons are carried out.

2.4.4. Relationships between Possible Predictor Variables and Arithmetic

After exploring the relationship between the comparison tasks the next step was to investigate the relationships between the comparison measures and arithmetic achievement. Arithmetic was defined as a measure that assesses basic number skills (e.g. identifying and writing numbers and counting) and basic calculations involving addition, subtraction, multiplication and division. Two measures of arithmetic were chosen for this study: the first was a standardised test that assesses children’s ability to complete numerical operations (WIAT: Numerical Operations). The measure begins with identifying and writing numbers, counting items and then on to addition, subtraction and multiplication problems; these progress from simple one digit sums to two digit sums (see Appendix 1 for a list of the items). On this measure the children were under no time pressure to complete the items. The second arithmetic measure was one minute addition, which requires children to complete simple one and two digit addition problems. This test was a speeded measure and therefore assesses fluency at completing the sums. The relationship between the comparison tasks and these two arithmetic tests were analysed separately, this allows for the investigation of whether any relationships found are related to arithmetic achievement in general (speeded and untimed), or perhaps just arithmetic fluency. The best fitting comparison task CFA was taken (magnitude comparison and letter comparison) and a further CFA was conducted including arithmetic and the control measures that might influence the relationship between children's magnitude comparison ability and arithmetic achievement (age, verbal ability, nonverbal ability, number identification). There were significant correlations between both the WIAT and speeded addition measures and both comparison latent variables, even with the other variables controlled for (age, verbal ability, and nonverbal ability). This finding provides support for previous studies who found a significant relationship between arithmetic achievement and performance on magnitude comparison tasks (e.g. symbolic: Durand et al. 2005; nonsymbolic: Libertus et al. 2011; both: Mundy & Gilmore, 2009).

Due to the large sample size it was more likely that any associations would be significant, therefore it is important to look at the strength of the associations. Taking WIAT arithmetic achievement, the correlations with both the magnitude and letter comparison tasks were moderate in strength (the association between arithmetic and magnitude comparison was slightly stronger). There were also significant relationships between children's arithmetic achievement and age, vocabulary, nonverbal ability, and number identification. The weakest association was between arithmetic and age; this is possibly due to the focussed age range of the children in the study. An interesting finding was that the strongest association was with number identification, which warranted further investigation. For speeded addition, there were again moderate associations with both comparison tasks but they were very similar in strength. The association with number identification was this time similar in strength to the comparison factors. As moderate relationships were found (which were stronger than some found by previous studies) this merited the investigation of the predictive relationships of the measures.

Correlations only inform us about the strength of the linear relationship between the measures and do not mean that an improvement on one measure will result in an improvement in the other (i.e. predict individual differences). Therefore to explore these relationships further a full path model was run to investigate which, if any, of the measures would predict significant variance in children's arithmetic achievement. Overall it was found that the variables explained 54% of variance in children's WIAT achievement and 40% of the variance in children's addition fluency scores. Magnitude comparison was found to be a significant predictor of children's arithmetic achievement (both untimed and speeded), this provides support for previous studies (e.g. symbolic: Durand et al., 2005; nonsymbolic: Libertus et al., 2011) and extends previous findings as the current study included controls for a variety of other factors. For example, the comparison process itself was controlled for due to the inclusion of the letter comparison task, children's general cognitive ability (both nonverbal ability and vocabulary) and age were also included in the CFA. However, these results are in contrast to previous studies that found that individual differences on symbolic but not nonsymbolic comparison tasks predicted variance in children's arithmetic scores (e.g. Holloway & Ansari, 2009; Sasanguie, De Smedt et al., 2012), and studies that found significant differences between children with arithmetic scores in the average range and those described as having a mathematics learning disorder on symbolic but not nonsymbolic comparison tasks (e.g. De Smedt & Gilmore, 2011). Although letter comparison was found to be related to children's arithmetic achievement it did not predict significant variance in their arithmetic scores, suggesting that whilst digits and letters have an ordinal structure, it is the action of comparing the magnitude information that is represented by digits that predicts variance in children's arithmetic skill (alongside the comparison of nonsymbolic numerosities).

Although the inclusion of a measure of children's knowledge and familiarity of numbers (number identification) was initially included to act as a control measure it was found to be a significant predictor of variance in children's arithmetic scores (alongside magnitude comparison). This may be intuitive that children who have greater knowledge of symbolic numbers are also more proficient on tests which use the symbolic number system; however, there is conflicting evidence on this in the wider literature. This finding that familiarity with symbolic numbers is important for arithmetic achievement is consistent with existing literature with both typically developing children and children with mathematical difficulties (e.g. De Smedt & Gilmore, 2011; Gilmore et al., 2010; Landerl et al., 2004; Lembke & Foegen, 2009; Mazzocco & Thompson, 2005). Nevertheless, it is in contrast to some evidence from typically developing children that performance on number recognition (reading) tasks is not related to arithmetic achievement (De Smedt et al., 2009; Soltész, Szűcs & Szűcs, 2010; however see Gilmore et al., 2010). The number identification measure used in the present study required children to match a verbal number word to its visual Arabic form, ignoring the distractor items that included common errors that young children make; this task is therefore more complex than asking children to simply read numbers that they are highly familiar with (i.e. 1 to 10). The task is potentially assessing children's knowledge of the symbolic system, knowledge of what the symbol represents (i.e. mapping the symbol to the magnitude it represents) and/or knowledge of place value. All of these skills could be important for arithmetic achievement. Children who understand place value may also be better at checking their answers on the arithmetic task and correcting any errors. This predictive relationship observed between number knowledge and arithmetic could be analogous to the relationship between letter knowledge and reading ability.

Neither age nor general ability (nonverbal ability or vocabulary) were predictors of variance in children's WIAT arithmetic scores. However, age was a significant predictor of variance in children's performance on the speeded arithmetic task. The finding that differences in the age of the children did not predict individual differences in untimed arithmetic achievement is not surprising due to the small age range of the children. All children were from the same year group at school so will have received a similar degree of formal numeracy teaching. The predictive relationship found between age and children's performance on the speeded arithmetic measure could therefore reflect speed of processing differences between the youngest and oldest children (e.g. Kail, 1991; Kiselev, Espy & Sheffield, 2009).

2.4.5. Future Research Questions

This current study has found that children's magnitude comparison performance and number identification ability significantly predict individual differences in arithmetic achievement in 5 to 6 year old children. However due to the concurrent nature of the study we are unable to establish the causal direction of the relationship. The question remains whether a good knowledge of the symbolic number system and efficient magnitude comparison (symbolic digits and nonsymbolic items) lead to better arithmetic achievement or whether ability/performance on these tasks are a consequence of learning arithmetic. The same children in this study will therefore be tested again on the same measures at later time points to investigate this (see Chapters 4 and 5).

The finding that number identification was a significant predictor of individual differences in children's arithmetic scores was surprising. Initially the task was included as a control measure for knowledge of the symbolic number system. Variation in number knowledge has been more typically explored in children with mathematics difficulties compared to children with scores in the average range (e.g. Landerl et al., 2004). The evidence for a relationship between performance on number recognition (reading) tasks and arithmetic achievement in typically developing children is inconsistent (e.g. De Smedt et al., 2009; Gilmore et al., 2010; Soltész et al., 2010). This therefore warrants further investigation and the question remains whether this measure predicts development of arithmetic skill. As mentioned previously, by assessing children's arithmetic achievement again at later time points, this will allow for the investigation of whether number identification is a longitudinal (as well as a concurrent) predictor of variance in arithmetic scores (see Chapters 4 and 5).

This study employed a novel presentation of the comparison tasks by using a group presented paper and pencil task rather than presenting the task to children individually on the computer, however the same effects (distance and ratio) were present in children's comparison of the items. Previous studies have shown that the effect of distance on comparison decreases with age (e.g. Holloway & Ansari, 2008; Sekuler & Mierkiewicz, 1977); however these studies include different children at different ages (i.e. are cross-sectional) rather than following the same children's development over time. Therefore the question remains over children's performance on these comparison tasks longitudinally. This question will be addresses in the following chapter (Chapter 3).

The relationships found were concurrent and with children early on in their formal schooling (aged 5 and 6); therefore from this study we do not know whether these associations are age specific. The relationships may differ depending on the age of the children and therefore their experience of arithmetic teaching, for example knowledge of the symbolic number system may be less important for arithmetic achievement when children are older and have mastered this skill. This also warrants further investigation.

Chapter 3. Children’s Development on Group Presented Comparison Tasks

3.1. Introduction

This study had one main aim: to explore the effects of number processing that are typically observed in the literature using a longitudinal design rather than a cross-sectional design that is typically seen in the literature

In Chapter 2, an effect of distance on comparison performance was observed in all three of the different categories of comparison task (digit, nonsymbolic and letter) with children comparing more items when the numerical distance between the items was far than when it was close. This replicates the classic distance effect results seen in the literature (e.g. Moyer & Landauer, 1967; Holloway & Ansari, 2009; Jou & Aldridge, 1999). Similar to the effect of distance an effect of the ratio between the two nonsymbolic numerosities was also observed on children’s comparison. As the ratio between the numerosities decreased (bigger number/smaller number) children’s performance decreased, which is in line with previous findings (e.g. Halberda & Feigenson, 2008). These effects are suggested to reflect an internal representation of number, in that numbers are represented approximately as magnitudes on an internal number line (also referred to as the approximate number system; Dehaene, 1992; Gallistel & Gelman, 1992). Thus the smaller the numerical difference (distance or ratio) between two numerosities the more difficult it is to discriminate between them as they would be represented closer together on the number line (and therefore the variability of the representations would have greater overlap) than would numerosities with a larger numerical difference.

As the distance and ratio effects are thought to be an indicator of the precision of an individual’s representation of number, these effects on children’s comparison performance were explored in Chapter 2. An effect of distance was found on children’s digit, letter and nonsymbolic comparison, as well as an effect of ratio on children’s comparison of larger numerosities. However, in that study a novel presentation method was used, with children completing paper and pencil versions of the tasks in a group testing setting (rather than computerised versions). The current chapter will retest the same children at three difference time points (and therefore different ages) and it is hypothesised that the findings will be replicated here with children comparing more digits, nonsymbolic numerosities and also letters when the distance between the two stimuli is larger (5, 6, or 7) than when it is smaller (1 or 2). It is expected that the distance effect will be present at each time point. It is also hypothesised that the effect of ratio on children’s comparisons will be replicated at each time point, with children gaining higher scores as the ratio between numerosities increases (7:8 > 5:6 > 3:4).

A second effect typically found in the literature is that overall performance on comparison tasks gets better with development. Accuracy at comparing the items has been found to increase with age (Halberda & Feigenson, 2008; Holloway & Ansari, 2008; Piazza et al., 2010), while time to compare the items (reaction time) has been found to decrease with increasing age (e.g. Budgen & Ansari, 2011; Holloway & Ansari, 2008; Sekuler & Mierkiewicz, 1977). These age related changes are suggested to reflect refinement in an individual's internal representation of number (Halberda & Feigenson, 2008). However, these studies include different children at different ages (i.e. are cross-sectional) rather than following the same children's development over time. In a recent study, Libertus et al. (2013) observed that over a six month period young children's (mean age at Time 1 was 4 years 2 months) accuracy on a nonsymbolic comparison task increased, while their reaction times decreased, although this change was not tested statistically. Reeve et al. (2012) assessed the same children seven times between the ages of 6 to 11 years; they found that with development children's reaction times at comparing symbolic numbers decreased. The current study aims to replicate these age related changes in comparison performance on the group presented comparison tasks. As an effect of age has been consistently found in the literature, especially over the early school years, it is hypothesised that children's performance on the group presented comparison tasks will increase with age.

As both age related changes in performance and effects of distance and ratio have been observed in children's performance on comparison tasks, the combination of these effects has been explored (i.e. the change in these effects over development). The effect of distance has been found to decrease with age when comparing both symbolic (e.g. Budgen & Ansari, 2011; Holloway & Ansari, 2008; Sekuler & Mierkiewicz, 1977; Sasanguie, De Smedt et al., 2012) and nonsymbolic (e.g. Holloway & Ansari, 2008; Sasanguie, De Smedt et al., 2012) stimuli. Children have also been found to get more accurate at comparing numerosities with increasingly difficult ratios with development (Halberda et al., 2008; Piazza et al., 2010). This change with increasing age is thought to represent the refinement of magnitude representations within this approximate number system (Halberda & Feigenson, 2008). However, these studies are cross-sectional rather than longitudinal and there is a paucity of longitudinal research into the development of these effects. Recently, Reeve et al. (2012) reported that the symbolic distance effect slope decreased between 6 and 11 years old, while Libertus et al. (2013) found that the size of the Weber fraction decreased over a six month time period (mean age at Time 1 was 4 years 2 months). Therefore the current study will add to this body of literature and investigate the development of the distance and ratio effects within the same sample of children over a two year period when the children will be, on average, 6, 7 and 8 years old. In the majority of previous studies participants complete comparison tasks individually on the computer, whereas this study employed a novel design and presented the tasks to children in the classroom setting as paper and pencil measures. This research will also add to the existing literature by examining whether these findings can be replicated in these relatively simple tasks. Taking the results from the previous studies it is hypothesised that the effect of distance and ratio will decrease over time.

The final aim of this chapter was to explore whether the relationships between the different types of comparison task changed over time, specifically to investigate the underlying latent factor structure of the comparison tasks at Time 2 and Time 3. In the previous chapter it was found that the best fitting model to the comparison data was a two factor solution, with the digit and nonsymbolic comparison tasks loading on the same factor and a separate letter comparison factor. This suggests that the digit and nonsymbolic comparison tasks are tapping the same underlying structure, comparison of numerical magnitude. This analysis will therefore be repeated for the Time 2 and Time 3 data to explore whether the same factor structure remains or whether it changes with development.

To explore these aims, the four different types of comparison task that were created for the study in Chapter 2 (digit, letter and the two nonsymbolic tasks) will be administered to the same group of children a further two times, with a time lag of almost one year between each time point. Typically, performance on magnitude (digit and nonsymbolic) comparison tasks has been investigated but in Chapter 2 a letter comparison task was included as a control measure for the comparison process. For completion development on this task and change in the distance effect will also be explored here. The difference between this current study and previous studies is that the comparison tasks are paper and pencil measures which will be presented to children in their class groups. They consequently assess both accuracy and speed at making comparisons (items correct within 30 seconds).

3.2. Method

3.2.1. Design

There were three time points of assessment. Children were assessed at time point 1 of the study between April and July 2010 (see Chapter 2 for more detail on the concurrent data). Time point 2 took place between March and June 2011 which was on average 11.74 months later (range = 10.83 to 13.27 months). The third time point took place in February and March 2012 with an average time lapse of 9.63 months after Time 2 (range = 9.33 to 11.00 months) and 21.37 months after the first testing phase (range = 20.20 to 22.90 months).

3.2.2. Participants

Details of the recruitment process and more information on the schools can be found in Chapter 2 (section 2.2.1/page 50). Children were recruited from four schools and at the first time point 173 children were tested from eight classes overall. At this time point all children were in school Year 1. At Time 2 (school Year 2) 165 children from the original sample were retested on average 12 months after the first testing phase. At time point 3 (school Year 3) 164 children of the original sample were retested a third time. The attrition rate of the sample was low with only eight children not being retested at Time 2; seven children had moved schools and one child was on holiday during the testing period. At Time 3 only nine children from the original sample were not retested as they had all moved schools. Where children joined a school after the first testing point their data were not included. At Time 1 children ranged in age from 5 years 8 months to 6 years 9 months, at Time 2 children ranged in age from 6 years 8 months to 7 years 8 months, while at Time 3 children now ranged from 7 years 5 months to 8 years 6 months old. Details for the sample can be found in Table 3.1.

Table 3.1.
Information on the group testing sample at all time points

	Time 1	Time 2	Time 3
N	173	165	164
Gender (male:female)	97:76	93:72	92:72
Mean age (*SD*) in months	74.69 (3.43)	86.26 (3.47)	95.79 (3.40)
Number of classes	8	8	9

3.2.3. Assessment Battery

At time points 2 and 3 children were retested on measures from the first time point and new tasks were also introduced. This study aims to explore children's development on the comparison tasks therefore only these tasks will be detailed here. Detailed descriptions of the other measures can be found in other chapters. The same comparison tasks that were administered at Time 1 were presented to the children again but the number of items was increased to avoid ceiling effects. A more detailed description of the construction and presentation of the tasks is presented in Chapter 2 (section 2.2.2.6, page 53), but a summary of the tasks is presented below.

At Time 1 children completed 14 comparison tasks, however five tasks were removed from the analysis due to a possible presentation order effect on the nonsymbolic ratio manipulated tasks and also to simplify the analyses (more detail on this can be found in Chapter 2). It was therefore decided at Time 2 and 3 to only present the tasks that were included in the Time 1 analyses. This also allowed for all comparison tasks to be presented in one booklet and to be completed within the same testing session. For the full list of the comparison tasks presented at Time 1 see Appendix 4 and for Time 2 and 3 see Appendix 13.

The comparison tasks were presented as paper and pencil measures with each different comparison task (e.g. digit far, letter close) presented individually within an A5 booklet. Comparison items were presented in pairs, one on the left and one on the right hand side, with six pairs per page. The items were presented in a pseudo random order with the restriction that the same item never appeared directly underneath itself. Table 3.2 gives a summary of the comparison tasks presented to the children and includes the short hand labels used to refer to the individual tasks in the results section. The tasks were presented in the following order: letter close, digit far, NS 7:8 SA, NS close SS, letter far, digit close, NS far SS, NS 5:6 SA, NS 3:4 SA. For each comparison task children were given 30 seconds to complete as many comparisons as they could. For every correct comparison children were awarded one point, the maximum score achievable for each task at each time point is presented in Table 3.3. It is therefore important to keep in mind that these tasks assess both accuracy and speed at making comparisons.

Table 3.2.
Summary of group presented comparison tasks

Comparison task	Versions	Label	Item range	Distance
Symbolic digit	Far	Digit far	1 – 9	5, 6, 7
	Close	Digit close		1, 2
Letter	Far	Latter far	a – i	5, 6, 7
	Close	Letter close		1, 2
Nonsymbolic: Effect of distance	Far: Same size	NS far SS	5 - 13	5, 6, 7
	Close: Same size	NS close SS		1, 2
Nonsymbolic: Effect of Ratio	3:4: Surface area matched	NS 3:4 SA	20 - 40	n/a
	5:6: Surface area matched	NS 5:6 SA		n/a
	7:8: Surface area matched	NS 7:8 SA		n/a

Table 3.3.
Maximum score achievable on each group presented magnitude comparison task at each time point

	Time 1	Time 2	Time 3
Digit	36	48	60
Letter	36	36	48
NS Distance	36	48	60
NS Ratio	36	36	48

3.2.3.1. Symbolic digit comparison.

The items ranged from 1 to 9 and the pairs of digits were presented in both orders (5 and 3, 3 and 5). To allow for the investigation of the effect that distance had on comparison (numerical distance effect), two versions of the task were created: close (comparison items with a numerical distance of 1 and 2) and far (numerical distance of 5, 6, and 7).

3.2.3.2. Letter comparison.

A letter comparison task presented in the same format as the digit comparison task was constructed. The letters a to i were used with the letter ‘a’ replacing the digit 1, ‘b’ replaced 2 and so on. The letter comparison tasks also included two versions: close and far.

3.2.3.3. Nonsymbolic comparison.

All stimuli consisted of arrays of black squares within a black border on a white background. There were two versions of the nonsymbolic task, one to investigate the distance effect and a second to investigate the effect of ratio. These tasks varied in the size of the numerosities presented. The comparison items used to investigate the effect of distance used squares that were all the same size (SS) meaning that larger numerosities had more squares therefore a larger surface area of black. The two stimuli pairings used to investigate the effect of ratio were matched for total surface area (SA) therefore smaller numerosities had larger squares but the same surface area of black. By including stimuli that were some of the time the same size and the other times controlled for total surface area it was hoped that children would discriminate based on numerosity rather than nonnumerical cues. A detailed description of the construction of the stimuli can be found in Appendix 7.

3.2.3.3.1. Effect of distance.

This task was matched to the format of the digit and letter comparison tasks but to avoid the subitizing range the numerosities presented ranged from 5 to 13. The numerosity 5 replaced the digit 1, 6 replaced 2, and so on. Again there were two versions: close (numerical distance of 1 or 2) and far (numerical distance of 5, 6 or 7).

3.2.3.3.2. Effect of ratio.

This task was constructed to investigate the effect of the ratio between the two numerosities being compared. It included larger numerosities that ranged from 20 to 40. The ratios between the numerosities were 3:4, 5:6, and 7:8 and each of these tasks were presented separately. The numbers 20 through to 30 were used as a baseline and the nearest whole number at each ratio was calculated. For example 20 was compared to 27 (3:4), 24 (5:6), and 23 (7:8).

3.2.4. Procedure

Children were tested in whole class groups at each time point, these ranged from 13 to 30 children. The measures were presented within two of three sessions at Time 1 and one of two sessions at time points 2 and 3, sessions lasted on average an hour each. Other tasks were included at these time points but are not included here as they are not the focus of this chapter which explores the children’s development on the comparison tasks (for the full list of tasks presented to children at Time 1, 2 and 3 see Appendix 8, 11 and 12 respectively). Each child had their own booklet to mark their answers in. For each different comparison task children were presented with an example item with the correct answer 'ticked'. The concept of the task was explained (i.e. that they had to choose the bigger number, the box with the most squares, or the letter that was later in the alphabet). A second pair was presented below and children were then asked to tick the answer to the experimenter's request (e.g. the bigger number). The experimenter checked that the children understood what they were required to do and explained that they would be given 30 seconds to compare as many pairs as possible. Children were then taken through each comparison task.

3.3. Results

3.3.1. Descriptive Statistics for the Comparison Tasks

The descriptive statistics are presented first followed by the exploration of children's development on the comparison tasks. As children were group tested as whole classes some children were absent when the comparison tasks were administered and therefore some data is missing. Children's performance on the comparison tasks at each time point are presented in Table 3.4. Scores represent the number of items correct within the time limit of 30 seconds, it is therefore important to keep in mind that these tasks assess both accuracy and speed at making comparisons. There was a wide range of scores on all measures, although not all of the data fit parametric assumptions (normality tests). However, analysis of variance is a robust statistical procedure and violations of the assumptions often have minimal effects (Howell, 2010, pp. 334), therefore parametric statistics will be used.

Table 3.4.

Descriptive statistics for the group administered measures

		N	Max achievable score	Mean	SD	Min	Max
Time 1	Digit far	164	36	20.54	6.02	5	36 (2)
	Digit close	159	36	17.28	5.68	2	34
	Letter far	161	36	8.59	4.69	0 (4)	24
	Letter close	160	36	5.16	3.38	0 (10)	18
	NS far SS	160	36	21.91	6.68	1	36 (4)
	NS close SS	163	36	12.56	4.71	0 (2)	25
	NS 3:4 SA	160	36	13.17	5.02	0 (2)	28
	NS 5:6 SA	159	36	11.14	4.38	1	26
	NS 7:8 SA	164	36	8.73	3.67	0 (1)	20
Time 2	Digit far	153	48	25.14	5.16	9	38
	Digit close	153	48	20.97	4.61	12	36
	Letter far	153	36	15.42	5.91	0 (1)	35
	Letter close	153	36	7.87	3.69	1	18
	NS far SS	153	48	28.27	6.55	12	48 (1)
	NS close SS	153	48	17.16	4.98	1	29
	NS 3:4 SA	153	36	17.39	4.85	5	33
	NS 5:6 SA	153	36	14.45	3.85	5	25
	NS 7:8 SA	153	36	11.24	3.62	3	22
Time 3	Digit far	156	60	30.06	5.49	12	48
	Digit close	156	60	25.66	4.95	13	39
	Letter far	156	48	17.93	6.26	1	37
	Letter close	156	48	9.84	4.26	2	22
	NS far SS	156	60	34.41	7.81	9	60 (1)
	NS close SS	156	60	19.42	4.99	2	32
	NS 3:4 SA	155	48	20.55	5.00	9	36
	NS 5:6 SA	156	48	16.83	4.53	4	31
	NS 7:8 SA	156	48	12.38	3.91	3	22

Note. Far = distance of 5, 6, and 7, close = distance of 1 or 2, NS = nonsymbolic, SS = same size stimuli, SA = surface area controlled stimuli. Figures in parentheses represent the number of children scoring at floor and ceiling.

3.3.2. Statistical Analyses of the Comparison Tasks

3.3.2.1. Digit comparison.

From Figure 3.1 it can be seen that at both distances children's scores are increasing over time and that at each time point children gained higher scores when the digits were far apart than when they were close together. To investigate the effect of time and distance on children's digit comparison ability a repeated-measures ANOVA was performed, with both time (time points 1, 2, and 3) and distance (far and close) as within subject variables. Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of time, $\chi^2(2) = 32.25$, $p < .001$, and the interaction, $\chi^2(2) = 7.96$, $p = .019$. Therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .82$ for the main effect of time and $\varepsilon = .94$ for the interaction). There was a significant effect of time on children's comparisons, $F(1.63, 207.21) = 253.42$, $p < .001$, $\eta_p^2 = .67$. Within-subjects contrasts revealed that children gained higher scores at Time 2 compared to Time 1, $F(1, 127) = 94.19$, $p < .001$, $\eta_p^2 = .43$, and at Time 3 compared to Time 2, $F(1, 127) = 274.97$, $p < .001$, $\eta_p^2 = .68$. A significant effect of distance was also observed, $F(1, 127) = 218.69$, $p < .001$, $\eta_p^2 = .63$, which confirms that children gained higher scores when the digits were far apart than when they were close together.

The interaction between time and distance was also significant, $F(1.89, 239.34) = 5.89$, $p = .003$, $\eta_p^2 = .04$, indicating that the effect of distance was different over time. Within-subject contrasts compared each time point to the previous and the two distances. They revealed that when the effect of distance was compared at Time 1 and Time 2 there was a significant interaction, $F(1, 127) = 8.48$, $p = .004$, $\eta_p^2 = .06$. From Figure 3.1 we can see that the effect of distance between Time 1 and Time 2 has actually increased. However, there was no interaction when comparing the effect of distance at Time 2 and Time 3, $F(1, 127) = 0.04$, $p = .833$, $\eta_p^2 = .00$, which suggests that it has remained stable between the later two time points.

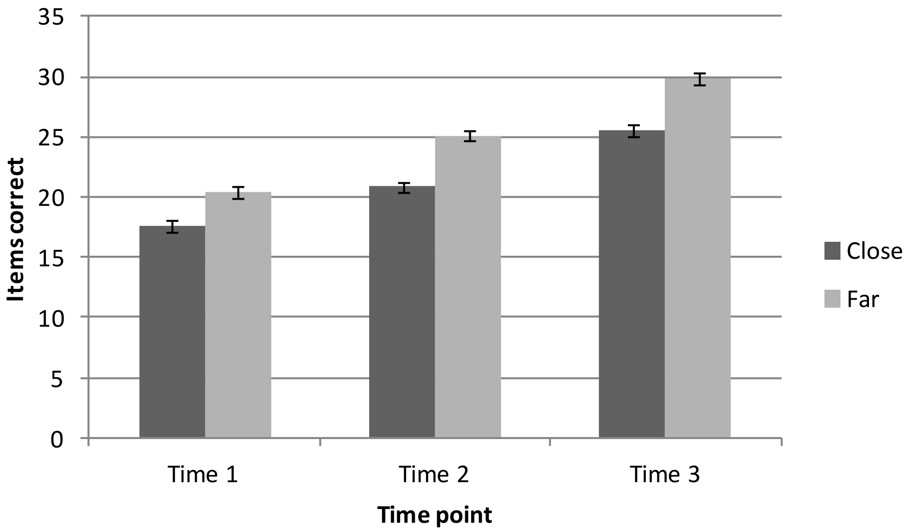


Figure 3.1. The effect of distance on children's performance on the digit comparison tasks over time. Error bars represent standard error.

3.3.2.2. Letter comparison.

The effect of distance on children's letter comparison ability over time was then explored. From Figure 3.2 it can be seen that at each time point children gained higher scores when the letters were far apart than when they were close together. At both the far and close distances children completed more items at Time 3, followed by Time 2, and gained the lowest scores at Time 1. It also appears that the effect of distance on children's comparison ability was smallest at Time 1. To investigate this statistically a two-way repeated-measures ANOVA was performed. The effect of time and distance was investigated (time: time points 1, 2, and 3; distance: far and close). Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of time, $\chi^2(2) = 10.70$, $p = .005$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .92$). A significant effect of time, $F(1.85, 230.91) = 203.38$, $p < .001$, $\eta_p^2 = .62$ was observed, within-subjects contrasts revealed that children gained higher scores at Time 2 compared to Time 1, $F(1, 125) = 191.47$, $p < .001$, $\eta_p^2 = .61$, and at Time 3 compared to Time 2, $F(1, 125) = 55.83$, $p < .001$, $\eta_p^2 = .31$. There was a significant effect of distance, $F(1, 125) = 661.75$, $p < .001$, $\eta_p^2 = .84$, which confirms that children were able to compare more items when the distance between them was far, than when it was close.

As was observed on the digit comparison task, the interaction between time and distance was also significant, $F(2, 250) = 40.21$, $p < .001$, $\eta_p^2 = .27$, indicating that the effect of distance was different over time. Within-subject contrasts revealed that when the effect of distance was compared at Time 1 and Time 2 there was a significant interaction, *F*(1, 125) = 51.13, $p < .001$, $\eta_p^2 = .29$. From Figure 3.2 it can be seen that the effect of distance increases. Children's scores on the close and far distances at Time 2 were compared to those at Time 3, the contrast was not significant, *F*(1, 125) = 2.12, $p = .148$, $\eta_p^2 = .02$. This suggests that the effect of distance remained the same between time points 2 and 3, this finding replicates that found when exploring the effect of distance on digit comparison performance over time.

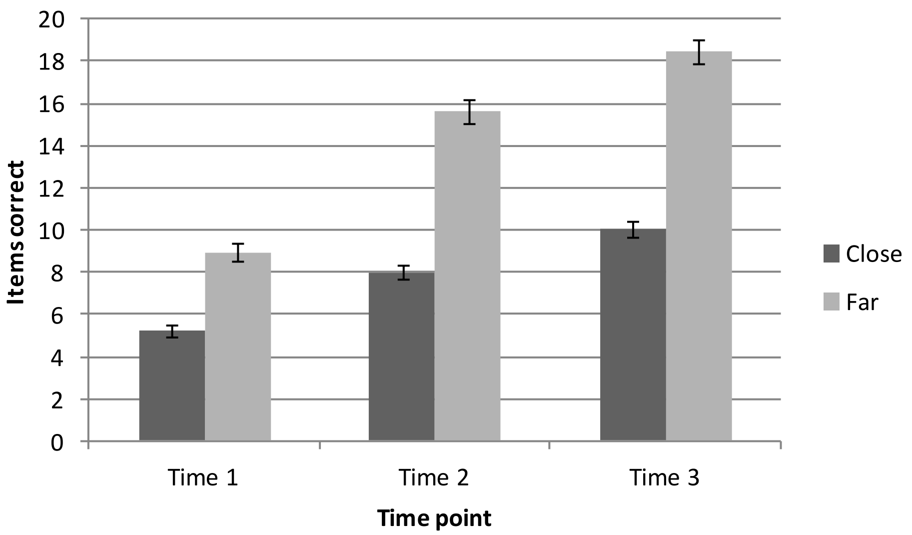


Figure 3.2. The effect of distance on children's performance on the letter comparison tasks over time. Error bars represent standard error.

3.3.2.3. Nonsymbolic comparison.

3.3.2.3.1. Effect of distance.

The effect of distance on children's nonsymbolic comparison ability was explored. From Figure 3.3 it can be seen that children's performance on the tasks increased over time and that at each time point children gained higher scores when the stimuli were far apart than when they were close together. It also appears that the effect of distance is increasing over time. To investigate this a 3 x 2 repeated-measures ANOVA was performed, with time point (time points 1, 2 and 3) and distance (far and close) as within subject variables. Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of time, $\chi^2(2) = 16.10$, $p < .001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .89$). There was a significant main effect of time on children's comparisons, *F*(1.78, 224.83) = 225.55, $p < .001$, $\eta_p^2 = .64$. Within-subjects contrasts revealed that children gained higher scores at Time 2 compared to Time

1, $F(1, 126) = 171.72$, $p < .001$, $\eta_p^2 = .58$, and at Time 3 compared to Time 2, $F(1, 126) = 104.24$, $p < .001$, $\eta_p^2 = .45$. There was also a significant effect of distance $F(1, 126) = 1447.01$, $p < .001$, $\eta_p^2 = .92$, with children gaining higher scores when the distance between items was far than when it was close. The interaction between time and distance was also significant, $F(2, 252) = 484.64$, $p <.001$, $\eta_p^2 = .21$, indicating that the effect of distance was different over time. Within-subject contrasts revealed that the difference in children's performance on the close and far distances at Time 1 compared to Time 2 was significant, $F(1, 126) = 5.28$, $p = .023$, $\eta_p^2 = .04$, as it was at Time 2 compared to Time 3, $F(1, 126) = 3.23$, $p < .001$, $\eta_p^2 = .22$. This indicates that the effect of distance is actually increasing between each time point.

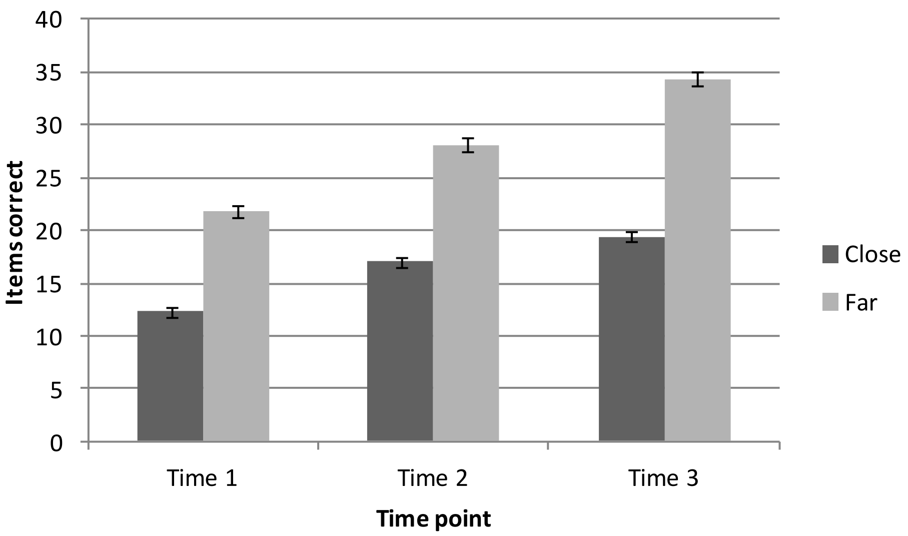


Figure 3.3. The effect of distance on children's performance on the nonsymbolic comparison tasks over time. Error bars represent standard error.

3.3.2.3.2. Effect of ratio.

The effect of ratio on children's nonsymbolic comparison ability was then explored. From Figure 3.4 it can be seen that within each time point children's scores decreased as the ratio between the stimuli became more difficult (number of items correct at 3:4 > number of items correct at 5:6 > number of items correct at 7:8). At each ratio children also compared more items correctly at Time 3, followed by Time 2, and gained the lowest scores at Time 1. It also appears that the effect of ratio may increase over time. To investigate this a 3 x 3 repeated-measures ANOVA was performed on the number of correct comparisons children made, with time point (1 to 3) and ratio (3:4, 5:6, and 7:8) as within subject variables. Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of time, $\chi^2(2) = 17.32$, $p < .001$, and ratio, $\chi^2(2) = 20.21$, $p < .001$. Therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε = .88 for the main effect of time and ε = .87 for the main effect of ratio). There was a significant main effect of time on children's comparisons, $F(1.77, 221.16) = 147.81$, $p < .001$, $\eta_p^2 = .54$. Contrasts revealed that children gained higher scores at Time 2 compared to Time 1, $F(1, 125) = 123.88$, $p < .001$, $\eta_p^2 = .50$, and at Time 3 compared to Time 2, $F(1, 125) = 61.64$, $p < .001$, $\eta_p^2 = .33$. A significant main effect of ratio, $F(1.74, 217.31) = 382.62$, $p < .001$, $\eta_p^2 = .75$, was also found. Within-subjects contrasts revealed that children gained higher scores when comparing items at the 3:4 ratio than the 5:6 ratio, $F(1, 125) = 213.64$, $p < .001$, $\eta_p^2 = .63$, and when comparing items at the 5:6 ratio than the 7:8 ratio, $F(1, 125) = 262.95$, $p < .001$, $\eta_p^2 = .68$.

The interaction between time and ratio was also significant, $F(4, 500) = 13.80$, $p <.001$, $\eta_p^2 = .10$, indicating that the effect of ratio differed over time. Within-subject contrasts compared each time point to the one before and each ratio to the next (i.e. compared 3:4 with 5:6, and 5:6 with 7:8). There was no change in the effect of ratio when comparing children's scores at Time 1 to Time 2 on the 3:4 and 5:6 ratios, $F(1, 125) = 2.63$, $p = .107$, $\eta_p^2 = .02$, and the 5:6 and 7:8 ratios $F(1, 125) = 0.25$, $p = .619$, $\eta_p^2 = .00$. When contrasting Time 2 and Time 3 the difference between children's performance on the 3:4 and 5:6 ratios again remained the same, $F(1, 125) = 3.11$, $p = .080$, $\eta_p^2 = .02$. However, there was a significant contrast when comparing performance on the 5:6 and 7:8 ratios, $F(1, 125) = 13.08$, $p < .001$, $\eta_p^2 = .10$. This suggests that the effect of ratio remained the same over time on the 3:4 and 5:6 ratios but there was a change in the ratio effect when comparing performance on the 5:6 and 7:8 ratios but only between Time 2 and Time 3.

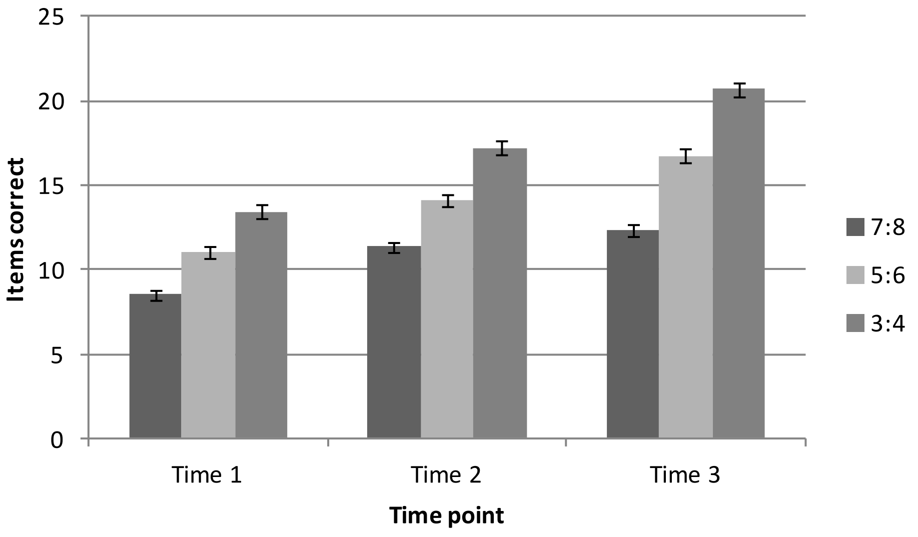


Figure 3.4. The effect of ratio on children's performance on the nonsymbolic comparison tasks over time. Error bars represent standard error.

3.3.2.4. CFA of comparison measures.

In Chapter 2 a series of confirmatory factor analyses were carried out to assess the relationships between the comparison tasks at the first time point. For completeness that analyses will be repeated here for the Time 2 and Time 3 data.

3.3.2.4.1. Time 2.

First the two factor solution, magnitude (digit and nonsymbolic) comparison and letter comparison, that best fitted the data at Time 1 was explored for the Time 2 data. The model, presented in Figure 3.5, again provided an acceptable fit to the data, χ^2 (26) = 47.918, *p* = .006, *RMSEA* = .074 (90% *CI* = .040-.107), *CFI* = .971, *SRMR* = .038. The other proposed solutions: one factor, two factor (symbol and nonsymbolic) and three factor, were also carried out to explore whether any would provide a better fit to the data at this time point.

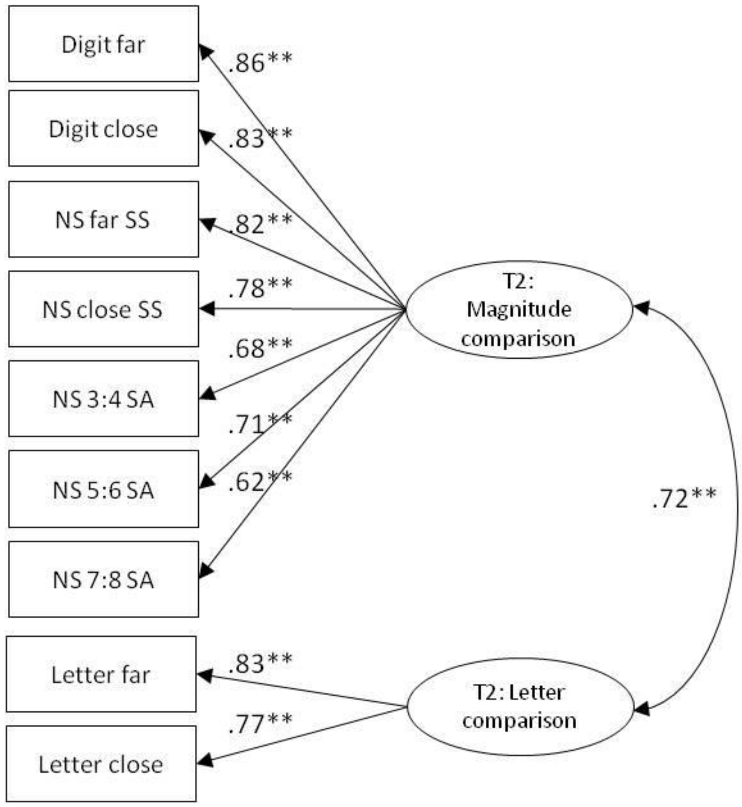


Figure 3.5. Two factor CFA of comparison tasks: Magnitude comparison and letter comparison (Time 2). Model fit: χ^2 (26) = 47.918, *p* = .006, *RMSEA* = .074 (90% *CI* = .040-.107), *CFI* = .971, *SRMR* = .038.
* *p* < .05, ** *p* < .01.

A one factor solution was explored to investigate whether all comparison tasks (digit, nonsymbolic and letter) represent one latent factor. Figure 3.6 shows the one factor CFA, which does not provide an adequate fit to the data, χ^2 (27) = 82.701, *p* <.001, *RMSEA* = .116 (90% *CI* = .088-.145), *CFI* = .927, *SRMR* = .054. The alternative two factor solution, with a symbol comparison construct (both digit and letter) and a separate nonsymbolic factor, was then performed. Again this CFA did not provide an acceptable fit to the data, χ^2 (26) =76.973, *p* < .001, *RMSEA* = .113 (90% *CI* = .084-.143), *CFI* = .933, *SRMR* = .051 (see Figure 3.7). The relationship between the two constructs was also very high (*r* = .95). The final possible structure of the comparison tasks is a three factor solution with each comparison task forming its own latent variable (digit, letter and nonsymbolic). This solution (see Figure 3.8) did provide an acceptable fit to the data, χ^2 (24) = 43.628, *p* = .008, *RMSEA* = .073 (90% *CI* = .037-.107), *CFI* = .974, *SRMR* = .036, however the coefficient between the digit and nonsymbolic latent variables was again high (*r* = .97).

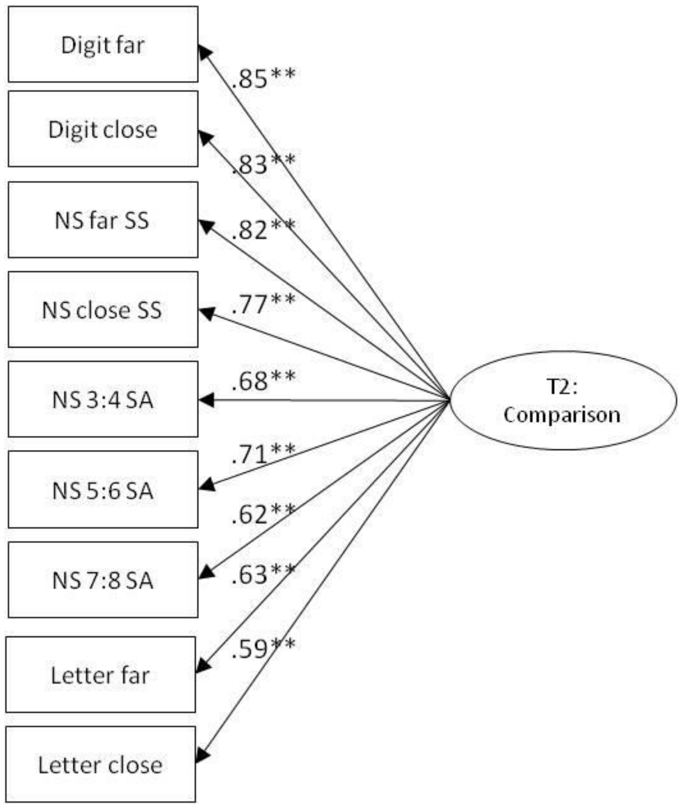


Figure 3.6. One factor CFA of comparison tasks (Time 2). Model fit: χ^2 (27) = 82.701, p <.001, *RMSEA* = .116 (90% *CI* = .088-.145), *CFI* = .927, *SRMR* = .054.

* p < .05, ** p < .01.

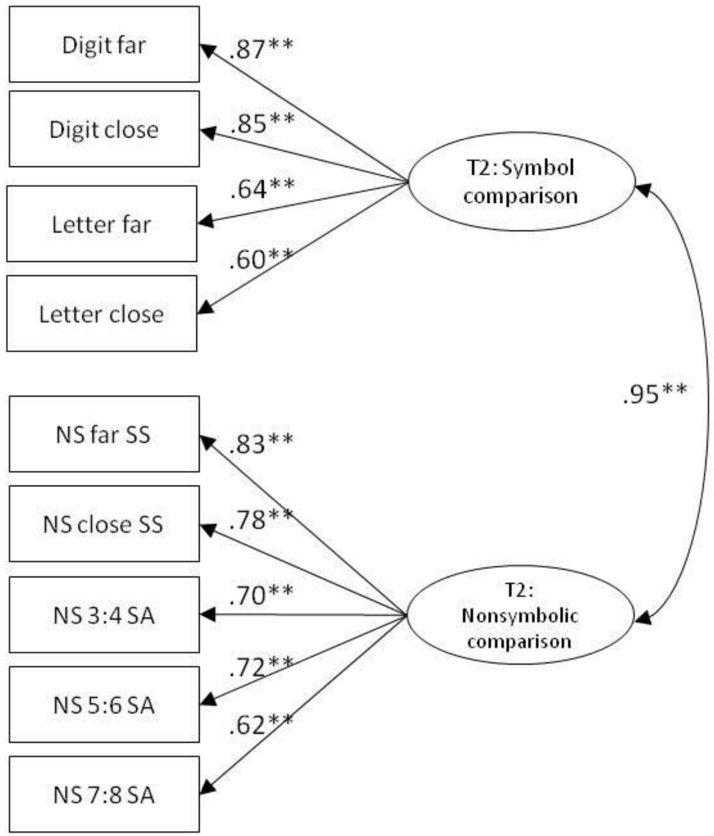


Figure 3.7. Two factor CFA of comparison tasks: Symbol comparison and nonsymbolic comparison (Time 2). Model fit: χ^2 (26) = 76.973, p < .001, *RMSEA* = .113 (90% *CI* = .084-.143), *CFI* = .933, *SRMR* = .051.

* p < .05, ** p < .01.

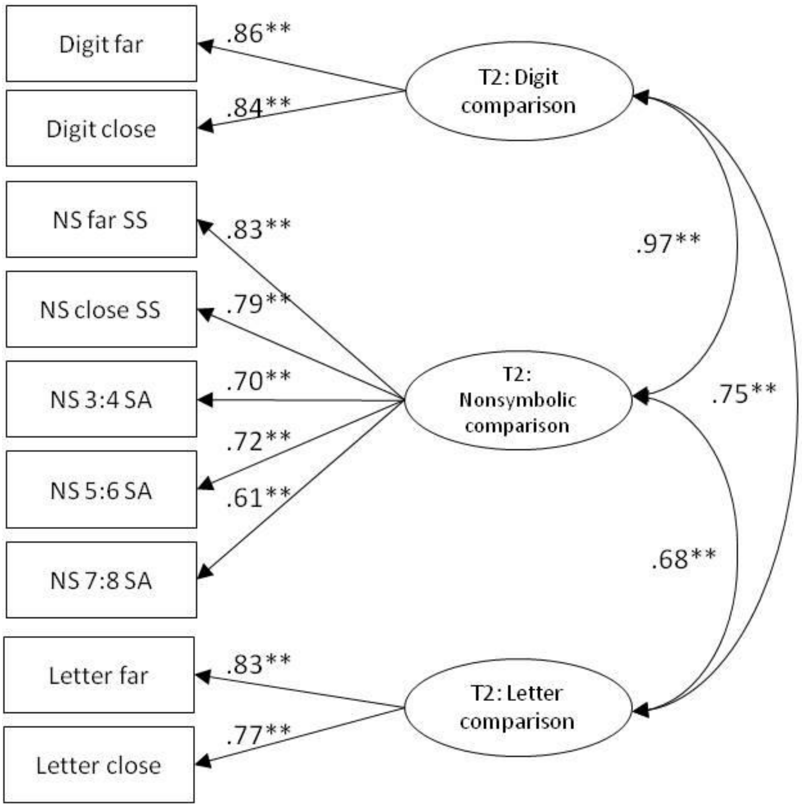


Figure 3.8. Three factor CFA of comparison tasks (Time 2). Model fit: χ^2 (24) = 43.628, *p* = .008, *RMSEA* = .073 (90% *CI* = .037-.107), *CFI* = .974, *SRMR* = .03.
* *p* < .05, ** *p* < .01.

Referring back to the model fit indices suggested as a guideline earlier, the two factor solution with magnitude (digit and nonsymbolic) and letter comparison latent factors and the three factor solution where every type of comparison task forms its own factor are the best fitting model solutions, and are both adequate fitting models. Both models fit the data equally well and a chi-square difference test revealed that they were not significantly different ($\Delta\chi^2$ = 4.64, df = 2, *p* > .05). However, in the three factor solution the relationship between the symbolic digit and nonsymbolic factors is *r* = .97, which suggests that these are not separate underlying constructs and are in fact the measuring the same underlying skill at this time point also. This finding replicates that found at Time 1 where the best fitting model is one where digit and nonsymbolic comparison tasks load on the same factor (magnitude comparison) whereas letter comparison forms a separate factor. There was a significant (moderate) relationship between the two latent variables which could reflect the method variance of the measures.

3.3.2.4.2. Time 3.

The analyses were then repeated for the Time 3 comparison data. First the two factor solution, magnitude (digit and nonsymbolic) comparison and letter comparison, that provided the best fit to the data at Time 1 and Time 2 was explored. The model, presented in *Figure 3.9*, did not provide an acceptable fit to the data at this time point, χ^2 (26) = 87.529, *p* < .001, *RMSEA* = .123 (90% *CI* = .095-.152), *CFI* = .926, *SRMR* = .043. The other proposed solutions were then run to examine whether any would provide a better fit to the data at this time point: one factor, two factors: symbol and nonsymbolic and the three factor solution.

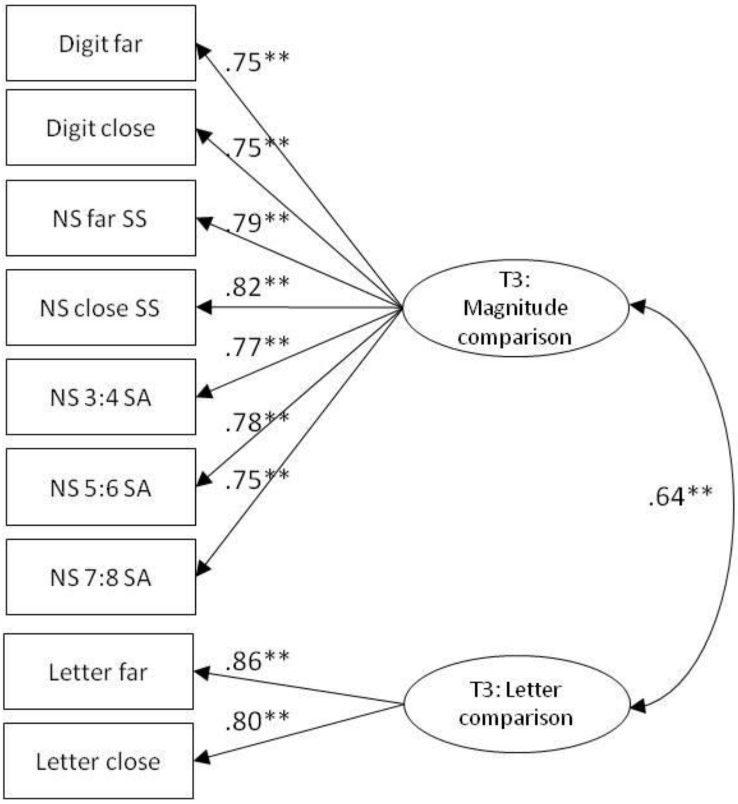


Figure 3.9. Two factor CFA of comparison tasks: Magnitude comparison and letter comparison (Time 3). Model fit: χ^2 (26) = 87.529, *p* < .001, *RMSEA* = .123 (90% *CI* = .095-.152), *CFI* = .926, *SRMR* = .043.
* *p* < .05, ** *p* < .01.

As found at Time 1 and Time 2 the one factor solution did not provide an adequate fit to the data, χ^2 (27) = 144.102, *p* <.001, *RMSEA* = .167 (90% *CI* = .141-.194), *CFI* = .858, *SRMR* = .067 (see *Figure 3.10*). The alternative two factor solution, with a symbol comparison construct (both digit and letter) and a separate nonsymbolic factor, was then performed. Again this CFA did not provide an acceptable fit to the data, χ^2 (26) =106.582, *p*

< .001, *RMSEA* = .142 (90% *CI* = .0115-.170), *CFI* = .901, *SRMR* = .055 (see *Figure 3.11*). The final possible structure of the comparison tasks is the three factor solution with each comparison task forming its own latent variable (digit, letter and nonsymbolic; see *Figure 3.12*). This solution was also less than an ideal fit to the data, χ^2 (24) = 56.272, *p* < .001, *RMSEA* = .093 (90% *CI* = .061-.125), *CFI* = .961, *SRMR* = .032.

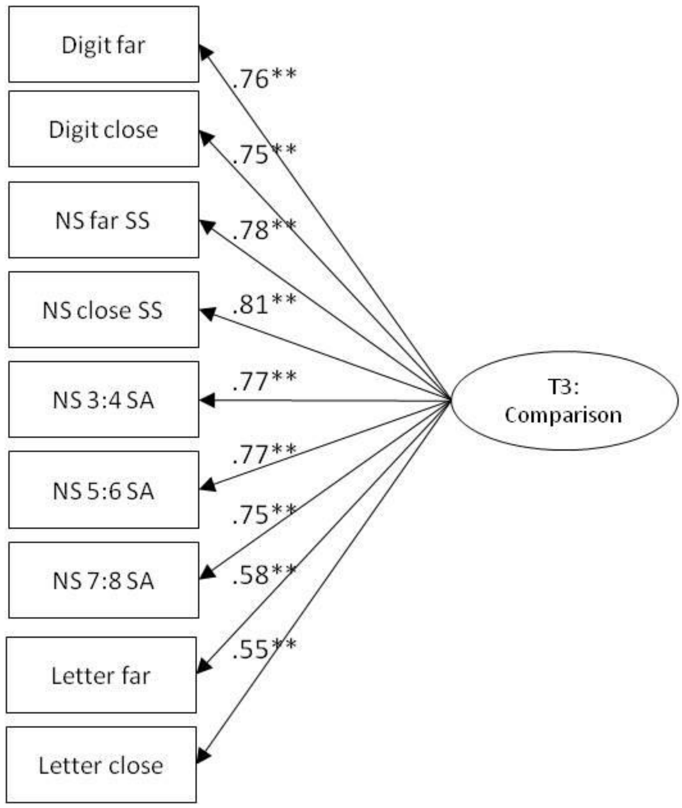


Figure 3.10. One factor CFA of comparison tasks (Time 3). χ^2 (27) = 144.102, *p* <.001, *RMSEA* = .167 (90% *CI* = .141-.194), *CFI* = .858, *SRMR* = .067.
* *p* < .05, ** *p* < .01.

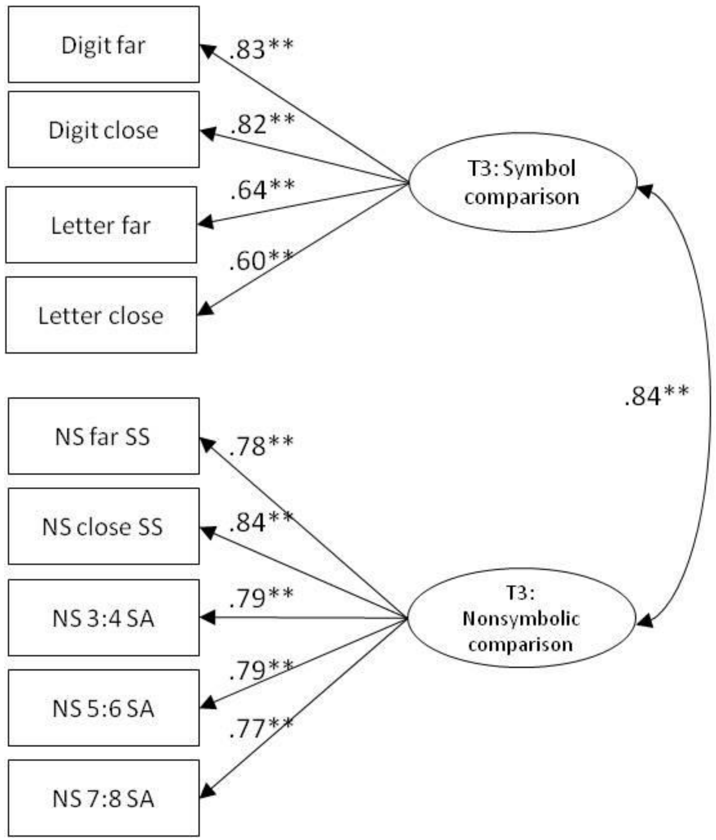


Figure 3.11. Two factor CFA of comparison tasks: Symbol comparison and nonsymbolic comparison (Time 3). χ^2 (26) =106.582, p < .001, *RMSEA* = .142 (90% *CI* = .0115-.170), *CFI* = .901, *SRMR* = .055.
* p < .05, ** p < .01.

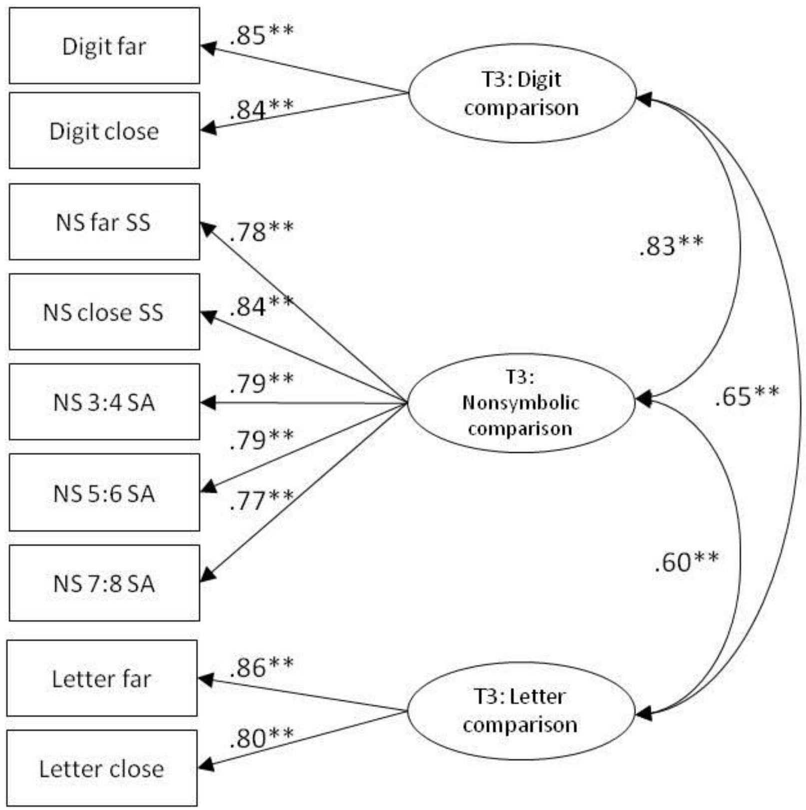


Figure 3.12. Three factor CFA of comparison tasks (Time 3). χ^2 (24) = 56.272, p < .001, *RMSEA* = .093 (90% *CI* = .061-.125), *CFI* = .961, *SRMR* = .032.
* p < .05, ** p < .01.

At this final time point none of the models provide an adequate fit to the data. Referring back to the model fit indices suggested earlier as a guideline, out of all the models the closest to an adequate fit is the three factor solution (Figure 3.12). The RMSEA value was not less than .08 (it was .09) but the CFI value was greater than .95 and the SRMR value was less than .05 (the latter two fit indices suggest that the model fit to the data is acceptable). In comparison the goodness of fit statistics of the other CFA's were outside of the suggested acceptable values. This suggests that at Time 3 the results differ to those found at the earlier time points, with the best fitting model being one with each comparison task forming its own latent variable, and suggests that these factors are separate underlying constructs. However, caution must be taken due to the less than ideal goodness of fit of the model to the data.

3.4. Discussion

The main aim of this study was to investigate the developmental effects that are typically observed in the literature using a longitudinal rather than a cross-sectional design. The comparison tasks used were also novel in that they were group presented paper and pencil comparison tasks.

Many studies have found that children's performance on magnitude comparison tasks improves with age, with accuracy increasing (e.g. Halberda & Feigenson, 2008; Holloway & Ansari, 2008; Libertus et al., 2013) and comparison times (RTs) decreasing (e.g. Budgen & Ansari, 2011; Holloway & Ansari, 2008; Reeve et al., 2012; Sekuler & Mierkiewicz, 1977). However, the majority of these studies are cross-sectional rather than longitudinal. This current study replicated the improvement over development in the same group of children, with children's scores (measured as number of items correct within 30 seconds) increasing on all of the comparison tasks (digit, nonsymbolic and letter) over a two year time period. In fact, on each type of comparison task children's performance increased with each time point (Time 1<Time 2<Time 3). These results provide support for those observed in the recent longitudinal studies (Libertus et al., 2013; Reeve et al., 2012), but also adds to the literature in that these were relatively simple group presented paper and pencil comparison tasks that were easy to administer and score compared to the individually presented computerised tasks. The highly replicated distance effect was also found on children's comparison performance, supporting the findings reported in Chapter 2 and in existing literature (distance: Moyer & Landauer, 1967; Holloway & Ansari, 2009; Jou & Aldridge, 1999). An effect of ratio on children's comparison of larger nonsymbolic

numerosities was also observed, again provided support for the findings in Chapter 2 and existing literature (Halberda et al., 2008; Sasanguie et al., 2013).

The current study also explored the age related changes of the distance effect in the same group of children. Both cross-sectional and more recently longitudinal, studies have shown that the effect of distance decreases with age, whether comparing symbolic or nonsymbolic stimuli (e.g. concurrent: Holloway & Ansari, 2008; Sasanguie, De Smedt et al., 2012; Sekuler & Mierkiewicz, 1977; longitudinal: Libertus et al., 2013; Reeve et al., 2012). The current study found that on the digit (symbolic) comparison task the size of the distance effect increased between Time 1 and Time 2 but then remained the same between Time 2 and Time 3, whereas on the nonsymbolic version the effect of distance was found to increase between each time point. There are possible reasons for these apparent contradictory findings, the first being the method used in this study to assess children's comparison ability. It differed to that typically seen in the literature as children were tested in whole class groups using a paper and pencil measure, rather than individually using a computer. The metric obtained from this study was the number of items children could compare correctly within 30 seconds, which encapsulated both speed and accuracy rather than just the accuracy rate or reaction times for comparing a set number of items. Therefore, over development children's ability to compare items that were far apart increased at a faster rate compared to their ability to compare items that were closer in distance, resulting in the increase in the distance effect.

The increase in the symbolic distance effect from Time 1 to Time 2 could also be due to the age of the children at these time points. At the first time point children were on average 6 years 3 months old while at Time 2 they were now 7 years 2 months old. It has been reported that young children do not automatically access the semantics of Arabic numerals (Girelli, Lucangeli & Butterworth, 2000). Girelli et al. (2000) found that when comparing numbers on the physical dimension (e.g. which is physically larger, 2 vs. 6) children at 6 years of age did not show an interference effect when trials were incongruent (i.e. their responses were not slower when there was a mismatch between numerosity and physical size). This would suggest that at the first time point the access to the underlying magnitude representation of the digits is less automatic so children are slower at comparing digits whether they are numerically far or close in distance. By the later time points children have developed this automatic access and therefore find it much easier to compare the digits when they are far apart, resulting in an increased distance effect.

On closer inspection of some of the results presented in the wider literature, the finding of a decrease in the distance effect is less clear. For example, Holloway and Ansari (2008) tested 6, 7 and 8 year old children and adults on symbolic, nonsymbolic, height and brightness comparison tasks. They reported that across all tasks the effect of distance decreased with age and that there was no interaction with task. However it is not clear whether the effect of distance decreased between each group or just between the youngest children and the adults. In addition to this, no age by distance interaction was found when analysing children's accuracy scores. Sasanguie, Van den Bussche and Reynvoet (2012) also investigated the change in the distance effect on symbolic and nonsymbolic comparison performance and tested children of similar age to the current study (mean ages = 5.6 years, 6.7 years and 7.6 years old). They calculated a regression slope to represent the distance effect on comparison speed and found that while the distance effect decreased on the nonsymbolic comparison task, there was no change in the symbolic distance effect. The change in the distance effect therefore may not be as clear cut as decreasing between successive ages.

With regards to children's performance on the letter comparison task, the effect of distance increased between Time 1 and 2, and then remained stable between Time 2 and 3. This is the same pattern of findings that was observed with children's digit comparison. Even though the children found the letter comparison task more difficult, the two tasks produced the same effect. As mentioned previously digits and letters are both over learned ordinal sequences (count numbers and alphabet string) and both sequences are suggested to be organised from left to right (e.g. Dehaene, Bossini & Giraux, 1993; Gevers, Reynvoet & Fias, 2003). The similarity of the effects in the two tasks may reflect this and the automatic access to the number and letter strings. However, the digits are at an advantage as they also have cardinal meaning, in that the digit also represents the number/quantity of items in a set. The access to this magnitude representation therefore may aid comparison, resulting in the much higher scores on the digit comparison task.

The current study also included a second nonsymbolic task which involved comparing larger numerosities where the ratio, rather than the distance between the arrays, was manipulated. Children's performance on this task was found to increase with development and as the ratio between the items increased (7:8 > 5:6 > 3:4), which is in line with previous research (e.g. Halberda & Feigenson, 2008; Halberda et al. 2008). The effect of ratio on children's comparison ability was found to remain stable between Time 1 and 2 but there was a change in the effect between Time 2 and 3, with the effect increasing when comparing numerosities separated by a 5:6 and 7:8 ratio. It appears that between Time 2 and Time 3, children's ability to compare two numerosities at the 3:4 and 5:6 ratio increased at a greater rate compared to their ability to compare two numerosities at a 7:8 ratio. This suggests that although children's ability to compare arrays that differ by a ratio of 7:8 has improved slightly over development the majority of children are still finding this task very difficult even when they are 8 years old. This is in line with Halberda and Feigenson's (2008) finding that children ranging in age from 6 to 8 years old were able to accurately discriminate arrays that differed by a 5:6 ratio.

Although not all of the findings in the present study replicated those found in the existing literature, the most important points to note are that an increase in children's magnitude comparison ability was found with increasing age. Alongside this, effects of distance and ratio were found on children's comparison of digits, nonsymbolic stimuli and letters. The difference in the methodology used (number of comparisons within 30 seconds) resulted in a greater increase (over time) in children's ability to discriminate between items that were numerically far apart than the increase in their ability to discriminate between numerically close items. However, when related to the proposed models of number representation (e.g. Dehaene, 1992; Gallistel & Gelman, 1992) the increase in children's scores over development suggests that their magnitude representations are in fact becoming more precise (Halberda & Feigenson, 2008; Sekuler & Mierkiewicz, 1977).

The final aim of this chapter was to explore whether the factor structure of the comparison tasks remains the same or whether it changes with development. At Time 1 it was found that the best fitting model to the comparison data was a two factor solution, with the digit and nonsymbolic comparison tasks loading on the same factor, and a separate letter comparison factor. This suggests that the digit and nonsymbolic comparison tasks are tapping the same underlying structure, magnitude comparison. The analysis was repeated for the Time 2 and Time 3 data. It was found that at Time 2, when children were 6 to 7 years old, the best fitting model to the data was again a two factor solution with the digit and nonsymbolic comparison tasks loading on the same factor, and a separate letter comparison factor. At the third time point, when children were 7 to 8 years old, none of the models provided an acceptable fit to the data but the best fitting of the models was the three factor solution with each of the different types of comparison task (digit, letter, nonsymbolic) loading on separate factors. There were still moderate to strong relationships between the different comparison tasks, with the strongest being between the digit and nonsymbolic factors (*r* = .83). As the fit of this CFA was less than ideal caution should be taken when interpreting the results, but this pattern suggests that at this older age, the processes involved in symbolic and nonsymbolic number comparison are becoming differentiated. Further research is therefore needed, perhaps including larger numbers of the digit and letter comparison tasks (as there were only two), to explore this change in the underlying factor structure.

3.4.1. Future research questions

The present study replicated the typical effects of distance and ratio seen in the existing literature using a group presented paper and pencil measure. The use of these measures may still be questioned, therefore to verify that these comparison tasks are assessing something similar to the computerised measures that are presented to children individually (i.e. the internal representation of number), then these two different methodologies need to be tested within the same group of children. Chapter 6 will therefore explore the nature of children's performance on these magnitude comparison tasks in a subgroup of children who took part in the current study. By administering a more tightly controlled computerised nonsymbolic comparison task this will enable the exploration of the change in children's numerical representations using a metric called the Weber fraction. This is an emerging area of research but what is lacking is a longitudinal design like that used in the current research.

A limitation of the current study is that no control for children's simple reaction time was included in the analyses. The changes seen over development could be due to age related changes in speed of processing (Kail, 1991; Kiselev, Espy & Sheffield, 2009). Therefore future research that contains a speeded element should also assess children's simple reaction time and then include this as a covariate.

This study was concerned with children's development of magnitude comparison, and in turn the development of the internal representation of number that these tasks aim to measure. As it has been proposed that these representations form the basis on which symbolic arithmetic builds upon then an important question is whether individual differences in these representations predict children's later arithmetic ability. The relationship between magnitude comparison and later arithmetic skill will now be explored in Chapters 4, 5 and 6.

Chapter 4.

Longitudinal Predictors of Children's Arithmetic Achievement

4.1. Introduction

In Chapter 2 the concurrent relationship between arithmetic and comparison ability was investigated. It was found that magnitude (digit and nonsymbolic) comparison performance, alongside children's ability to identify numbers, were significant predictors of individual differences in arithmetic achievement. However, due to the concurrent nature of the study the causal direction of the relationship could not be established. Therefore the question remains whether a good knowledge of the symbolic number system and efficient magnitude comparison (symbolic and nonsymbolic items) leads to better arithmetic achievement or whether ability/performance on these tasks are a consequence of learning arithmetic. To investigate this, the same group of children were followed up longitudinally and their arithmetic ability assessed on average one and two years after the initial time point.

It is important to identify predictors of later arithmetic achievement so that children who are at risk of later difficulties can be identified early using these measures and given intervention to try to ameliorate these difficulties. This has been successful in the area of children's literacy development (e.g. Muter, Hulme, Snowling & Stevenson, 2004; Hatcher, Hulme & Ellis, 1994) therefore it is hoped that the concurrent findings of the previous chapter can be extended longitudinally.

Table 4.1.

Table of published studies investigating the longitudinal relationship between arithmetic and magnitude comparison

Paper	Age (Y;M where possible)	N	Number range	Arithmetic/mathematics task	Classification	Correlation with arithmetic (r)	
						Accuracy	RTs
SYMBOLIC							
Desoete, Ceulemans, De Weerdt & Pieters (2012) - TD, MD, LA - Belgium	T1 = 5-6 T2 = 6-7 T3 = 7-8	315	TEDI-MATH Battery. Range = 2 to 6709	Kortrijk Arithmetic Test revision KRT-R (untimed) The arithmetic number facts test TTR (timed)	Arithmetic	No relationship between kindergarten comparison and Grade 1 arithmetic Kindergarten comparison with Grade 2 arithmetic r = .21** to .36**	
De Smedt, Verschaffel, & Ghesquière (2009) - Belgium	T1 = 6;04 (First grade) T2 = 7;03	T1 = 47 T2 = 42	1 to 9	Curriculum based standardised achievement test for mathematics	Mathematics	Overall accuracy r = .38*	Average RT r = -.45**, slope[a] r = .40*
Reeve, Reynolds, Humberstone & Butterworth (2012) - Australia	6 to 11 years	159	1 to 9	Single digit addition (6 years) Two digit computation (9.5 years) Multidigit computation test (11 years)	Arithmetic		The 3 groups differed significantly in their performance on the arithmetic measures at each time point (slow < medium < fast)
Sasanguie, Van den Bussche & Reynvoet (2012) - Belgium	Time 1 = 5.6yrs, 6.7yrs, 7.6yrs Time 2 = 1 year later	72	1 to 9	Curriculum based standardised achievement test for mathematics	Mathematics		RT (adjusted to reflect speed & accuracy) r = .31*, slope r = .08 (ns)

Sasanguie, Göbel, Moll, Smets & Reynvoet (2013) - Belgium	Time 1 = 6 to 8 yr olds (first to third grade) Time 2 = 1 year later	71	1 to 9	Timed arithmetic test (Tempo Test Rekenen[b])	Arithmetic		Median RT r = -.35** DE r = .07 (ns)
				Curriculum based standardised achievement test for mathematics	Mathematics		Median RT r = -.37** DE r = .06 (ns)
NONSYMBOLIC							
Desoete, Ceulemans, De Weerdt & Pieters (2012) - TD, MD, LA - Belgium	T1 = 5-6 T2 = 6-7 T3 = 7-8	315	TEDI-MATH Range = 1 to15	KRT-R (untimed) TTR (timed)	Arithmetic	Kindergarten comparison with Grade 1 arithmetic r = .16** to .20** The only significant relationship between kindergarten comparison and Grade 2 arithmetic was with number fact retrieval r = .16**	
Halberda, Mazzocco & Feigenson (2008) - retrospective	Weber fraction at 14yrs, Maths at kindergarten to 6th grade	64	5 to 16	TEMA-2 WJR- Calc	Arithmetic	WF was significantly correlated with previous arithmetic performance. Over the years ranged from r = .36** to .57**	
Libertus, Feigenson & Halberda (2013) - concurrent & longitudinal	Time 1 4;02 Time 2 4;09	Varies >100	4 to 15	TEMA-3	Arithmetic	% correct r = .46** WF r = -.25*	Mean RT r = -.33**

Mazocco, Feigenson & Halberda (2011a) - TD, MLD, LA, HA study[b] - Concurrent & longitudinal (retrospective)	Weber fraction: 14;10 (14;02 to 15;11) Arithmetic: Kindergarten to 6th grade	71	5 to 16 Ratios = 1:2, 3:4, 5:6, 7:8	TEMA-2 WJR-Calc	Arithmetic	WF was significantly correlated with arithmetic scores from all years. Ranged from r = -.32** to -.56**	
Mazzocco, Feigenson & Halberda (2011b) (these children were the youngest from Halberda & Feigenson, 2008)	Time 1: 4;02 (3;05 to 4;11) Time 2: 6;08 (6;02 to 7;5)	17	1 to 14	TEMA-3	Arithmetic	Regression analysis: % correct accounted for 28%* of variance while WF (n=14) accounted for 21% (ns)	
Sasanguie, Van den Bussche & Reynvoet (2012)	Time 1 = 5.6yrs, 6.7yrs, 7.6yrs Time 2 = 1 year later	72	1 to 9	Curriculum based standardised achievement test for mathematics	Mathematics		No relationships found
Sasanguie, Göbel, Moll, Smets & Reynvoet (2013)	Time 1 = 6 to 8 yr olds Time 2 = 1 year later	71	6 to 26 compared to reference numerosity of 16	Timed arithmetic test[b]	Arithmetic	No significant relationships using ether mean accuracy or WF	
				Curriculum based standardised achievement test for mathematics	Mathematics	No significant relationships using ether mean accuracy or WF	

Note. TD = typically developing, MLD = mathematical learning disability, LA = low mathematics achievement, HA = high achievement in mathematics. WF = Weber fraction. [a] Slope is an index of the effect of distance on comparison reaction times. [b] Tempo Test Rekenen; De Vos (1992)

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

There are fewer studies that have investigated the relationship between magnitude comparison and arithmetic longitudinally, than concurrently, and even less of those explore the contribution of both symbolic and nonsymbolic comparison performance in the same group of children (see Table 4.1). As with the cross-sectional literature, conflicting findings have been reported. De Smedt et al. (2009) found that the symbolic distance effect explained 10% of unique variance in children's mathematics achievement one year later (after controlling for age, nonverbal ability and number reading speed), whereas Sasanguie and colleagues found no association between the symbolic distance effect and either mathematics or arithmetic fluency (Sasanguie et al., 2013; Sasanguie, Van den Bussche & Reynvoet, 2012). They did however find that children's speed (RTs) on the symbolic comparison task was a significant predictor of children's later mathematics and arithmetic fluency (with age, spelling ability and other numerical processing measures controlled). Reeve et al. (2012) used both children's speed on a symbolic comparison task and the effect of distance and classified children as either slow, medium or fast; this group membership at 6 years old predicted children's arithmetic achievement at 9 and 11 years old. An interesting finding was reported by Desoete et al. (2012) who found that accuracy at comparing digits at 5 and 6 years old was not related to their arithmetic achievement one year later, but it was a significant predictor of their arithmetic achievement two years later (analysis only included the comparison tasks, i.e. no controls for age or cognitive ability).

Performance on nonsymbolic comparison tasks is suggested to represent the acuity of representations within the approximate number system (ANS). In general, it has been reported that accuracy at comparing nonsymbolic magnitudes is related to arithmetic achievement assessed at a different time (Halberda et al., 2008; Libertus et al., 2013; Mazzocco et al., 2011a, 2011b). However, two of the studies reported are retrospective (Halberda et al., 2008; Mazzocco et al., 2011a), with performance on the nonsymbolic comparison task assessed when individuals were 14 years old predicting individual differences in earlier arithmetic skill (5 to 11 years old). It could therefore be argued that greater acuity of the ANS is a consequence of learning arithmetic rather than superior comparison performance leading to better arithmetic achievement. On the other hand, the results from the prospective studies imply that later arithmetic achievement is predicted by children's prior ANS acuity (Libertus et al., 2013; Mazzocco et al., 2011b). However, two recent studies by Sasanguie and colleagues reported a lack of association between performance on a nonsymbolic comparison task and both arithmetic and mathematics achievement assessed one year later (Sasanguie et al., 2013; Sasanguie, Van den Bussche & Reynvoet, 2012).

Few studies have investigated both symbolic and nonsymbolic comparison performance together in relation to later arithmetic achievement. As reported above, Sasanguie and colleagues found that whilst children's speed at comparing symbolic numbers was related to later arithmetic and mathematics achievement, performance on a nonsymbolic comparison task was not (neither was the symbolic distance effect) (Sasanguie et al., 2013; Sasanguie, Van den Bussche & Reynvoet, 2012). Desoete et al. (2012) also included a nonsymbolic comparison task in their study and found an opposite pattern of findings to the symbolic comparison task; children's nonsymbolic comparison performance (assessed when they were 5 to 6 years old) was predictive of their arithmetic achievement one year later assessed by performance on untimed calculations, untimed word problems, and on a timed calculation fluency test. However two years later nonsymbolic comparison was only a significant predictor of their achievement on the timed calculation fluency task and did not predict achievement on the two untimed arithmetic measures. These differing results between symbolic and nonsymbolic comparison and also over the two time delay periods warrant further investigation. It should also be noted that the measures used by Desoete et al. differ to those typically used by other researchers (the tasks were taken from a battery designed to screen for Dyscalculia and included fewer items).

The findings reported for the role of symbolic and nonsymbolic processing ability for later arithmetic achievement are again inconsistent. As highlighted in the review of the literature in Chapter 1, and detailed in Table 4.1, there are many differences between these longitudinal studies. For example, most only explore either symbolic or nonsymbolic comparison as a possible predictor, while only a restricted few investigate the importance of both. The age ranges of the children differ with some studies including a narrow range, with others spanning multiple years. Many studies only investigate the predictive relationship one year later but it is possible that this may change with development. Further research is therefore needed that explores the longitudinal relationships between arithmetic achievement and both symbolic and nonsymbolic comparison performance together, over a one and two year time period and with a restricted age range of children.

In addition, a large number of the prospective longitudinal studies (De Smedt et al., 2009; Desoete et al., 2012; Mazzocco et al., 2011b) do not take into consideration children's prior arithmetic ability (the autoregressor). Although some of the children were quite young (some as young as 3 years 5 months) some of the children were 6 years old when they completed the magnitude comparison measures and will already possess some arithmetic knowledge. Therefore it is not clear whether this knowledge may have contributed to the relationship between magnitude comparison performance and arithmetic achievement (i.e. whether magnitude comparison performance is a unique predictor independent of earlier arithmetic knowledge). Sasanguie et al. (2013) carried out additional analyses controlling for children's mathematics achievement assessed at the same time as magnitude comparison ability and found that it explained 51% of variance in later mathematics scores. Furthermore children's symbolic comparison performance was no longer a significant predictor. Magnitude comparison ability may not explain additional variance in individual differences in children's arithmetic and mathematics beyond their prior knowledge. Further research is needed to clarify this.

In the current study, when analysing the predictors of children's arithmetic achievement, children's previous arithmetic ability will be controlled for in the analyses (i.e., controlling for autoregressive effects). Children's later arithmetic ability will depend, in part, on their earlier arithmetic skill (Gollob & Reichardt, 1987). By including the autoregressor in the analyses any relationships found between the predictor variables (age, general ability, comparison performance, number knowledge) and later arithmetic achievement are not influenced by children's earlier arithmetic knowledge (Hecht et al., 2001). Another advantage of including the autoregressor means that it is the *growth* in arithmetic ability from Time 1 to Time 2 that is being predicted (e.g., Gollob & Reichardt, 1987; Hecht et al., 2001). It has also been found that the relative contribution of a predictor variable can vary depending on whether the autoregressor is included in the analysis or not (see Gollob & Reichardt, 1987). In contrast, others have argued that including the autoregressor in analyses is problematic as a large amount of the variance to be predicted is potentially removed and that the autoregressive effect is in fact a stability coefficient (e.g. Chan, 2003). Taking this into consideration, and due to the fact that the majority of previous studies have not included children's prior arithmetic skill in the analyses, the predictors of children's later arithmetic will also be analysed without the autoregressor.

In Chapter 2 children's number identification ability, which was initially included as a control measure for knowledge of the symbolic number system, was found to predict individual differences in arithmetic scores. Previous research with children with mathematical difficulties, both concurrently (e.g. De Smedt & Gilmore, 2011; Landerl et al., 2004) and longitudinally (Mazzocco & Thompson, 2005), has found that children with lower mathematics ability perform more poorly on number reading tasks than children with mathematics ability in the average range. With regards to the typically developing literature both De Smedt et al. (2009) and Soltész et al. (2010) found that children's ability to read Arabic numerals was not related to their arithmetic achievement, which is inconsistent with the research into mathematical disorder. This may suggest that only children classified as having mathematics difficulties rather than just children with lower arithmetic ability have difficulty reading one digit numbers. It is also possible that the single digit number reading tasks are two easy for children that have begun their formal schooling and that there is reduced variance in the scores (i.e. children are at ceiling) which results in the lack of relationship with arithmetic in typically developing children.

When children's knowledge of Arabic numbers larger than ten has been investigated a relationship with their arithmetic achievement has been found. Gilmore, McCarthy & Spelke (2010) found that young children's Arabic number knowledge (assessed by naming one and two digit numbers ranging from 1 to 92) was significantly and strongly related to their mathematics ability assessed two months later (r = .71). Together with children's verbal number knowledge (assessed by the ability to decide which of two verbally presented one and two digit numbers was larger), number reading predicted arithmetic ability even when controlling for age, verbal ability and literacy achievement (11% of unique variance explained). In a study assessing measures for use as potential screening tools to identify children at risk of low mathematics achievement, Lembke and Foegen (2009) carried out a short longitudinal study with children in kindergarten and first grade. They included a number identification task (reading numbers ranging from 0 to 100) and found that performance on this task at the beginning of the school year was significantly related to formal arithmetic knowledge (TEMA-3) assessed at the end of the school year (kindergarteners r = .34, first grade r = .58). Taken together these studies suggest that assessing knowledge of the larger symbolic number system (i.e. beyond 10) may be a better predictor of children's arithmetic achievement than knowledge of the basic ten symbols (0 to 9). Longitudinal research carried out over an extended period of time, which includes a more difficult number identification measure and also controls for early arithmetic knowledge, is needed in order to clarify this.

Alternatively to assessing children's ability to read symbolic numbers (Arabic to verbal code), the ability to transcribe numbers presented verbally (verbal to Arabic code) has been investigated. Moeller et al. (2011) found that the overall number of errors children made when transcoding one, two and three digit numbers in first grade was significantly related to the number of errors they made on an addition task two years later (*r* = .24), although it was not a significant predictor. When considering only the number of inversion errors children made on the number writing task (e.g. writing 43 as 34) in place of total number of errors, this was a significant predictor of children's later addition competence. Children who made more inversion errors when writing digits made more errors on the addition task. It should be noted that the children involved in this study were German speaking so are more likely to make inversion errors due to the structure of the number words in the language. For example the number 43 is pronounced as dreiundvierzig (three and forty). Therefore it will be interesting to investigate this relationship in English speaking children.

In summary, the aim of the current study was to investigate the predictors of children's later arithmetic achievement. Children who completed potential predictor tasks early in their formal schooling (Year 1) again completed a standardised untimed arithmetic test (WIAT: Numerical Operations) almost one and two years later when they were in school Year 2 and Year 3. The CFA conducted in Chapter 2 was used as the measurement model to construct a path model predicting children's arithmetic achievement one and two years after the initial testing point. As it was found that letter comparison formed a separate latent factor to the magnitude comparison tasks, we could investigate whether the process of comparing two stimuli (general comparison performance) is related to arithmetic achievement, whether only magnitude comparison performance predicts later arithmetic skill, and the possibility that neither of the aforementioned predict individual differences in children's later arithmetic scores. The finding that children's number identification ability was a significant predictor of children's concurrent arithmetic achievement will also be investigated longitudinally. As in the previous chapter, the age of the children will be taken into account in the subsequent analysis, alongside their general cognitive ability. Finally, data will be analysed using sophisticated techniques, i.e. SEM, which will require the large sample of children to be retested.

4.1.1. Hypotheses

In Chapter 2 magnitude comparison ability, including both symbolic number and nonsymbolic comparison, was found to be a concurrent predictor of children's arithmetic achievement. It has also been reported in the wider literature that the efficiency with which symbolic numbers are compared is related to children's later arithmetic achievement. It is therefore feasible that magnitude comparison ability, again using a measure of both symbolic and nonsymbolic comparison, will be important for children's arithmetic development. In contrast to this, conflicting results have been reported for the importance of early nonsymbolic comparison ability for children's arithmetic development. In addition, nonsymbolic comparison ability, and with this the acuity of the approximate number system, has more typically been investigated using measures of accuracy performance rather than a metric of the speed with which comparisons are made (as symbolic comparison typically is). It is therefore also feasible that magnitude comparison ability may not be a predictor of later arithmetic achievement as the latent variable measure used here includes ability to compare both symbolic numbers and nonsymbolic numerosities. Few studies have controlled for prior arithmetic knowledge in the analysis and those that have, have reported differing results. Due to some weak correlations observed between magnitude comparison and arithmetic in the existing literature the inclusion and exclusion of the autoregressor may result in different findings regarding its importance for arithmetic development.

Due to the findings in Chapter 2 and in the wider literature that knowledge of the symbolic number system is important for arithmetic achievement it is hypothesised that children's number identification ability will be important for their later arithmetic achievement.

4.2. Method

4.2.1. Design

There were three time points of assessment. Children were assessed at time point 1 of the study between April and July 2010 (see Chapter 2 for concurrent data). Time point 2 took place between March and June 2011 which was on average 11.74 months later (range = 10.83 to 13.27 months). The third time point took place in February and March 2012 with an average time lapse of 9.63 months after Time 2 (range = 9.33 to 11.00 months) and 21.37 months after the first testing phase (range = 20.20 to 22.90 months).

4.2.2. Participants

The children included here were the same as those included in Chapters 2 and 3. At the first time point (school Year 1) 173 children were tested from eight classes overall. At Time 2, 165 children from the original sample were retested on average 12 months after the first testing phase. The final testing point, Time 3 (school Year 3), took place just under 12 months after the second testing phase and 164 children of the original sample were retested. The attrition rate of the sample was low with only eight children not being retested at Time 2; seven children had moved schools and one child was on holiday during the testing period. At Time 3 only nine children from the original sample were not retested as they had all moved schools. Where children joined a school after the first testing point their data were not included. At Time 1 children ranged in age from 5 years 8 months to 6 years 9 months, at Time 2 children ranged in age from 6 years 8 months to 7 years 8 months, while at Time 3 children now ranged from 7 years 5 months to 8 years 6 months. Details for the sample can be found in Table 4.2.

Table 4.2.
Information on the group testing sample at all time points

	Time 1	Time 2	Time 3
N	173	165	164
Gender (male:female)	97:76	93:72	92:72
Mean age (*SD*) in months	74.69 (3.43)	86.26 (3.47)	95.79 (3.40)
Number of classes	8	8	9

4.2.3. Assessment Battery

At time points 2 and 3 children were retested on measures from the first time point and new tasks were also introduced. This study is interested in how children's performance on the measures administered at Time 1 predict growth in arithmetic skill as assessed by the WIAT Numerical Operations subtest, therefore not all measures administered will be detailed here. The Time 1 variables that will be used to predict later arithmetic achievement are nonverbal ability, vocabulary, number identification, comparison performance and the autoregressor (WIAT Numerical Operations), a detailed description of these measures can be found in Chapter 2 (section 2.2.2, page 51). Regarding Time 2 and 3 tasks only the WIAT measure is relevant for this study so is detailed below.

4.2.3.1. WIAT-II: Numerical Operations subtest.

See Chapter 2 page 51 for a detailed description of the subtest used at Time 1. As the longitudinal study progressed and children advanced through the school curriculum additional items were added to the measure to avoid ceiling effects. For the Time 2 assessment a further three items were taken from the standardised test and presented to the children; these included a subtraction, multiplication and a division sum. Children were guided through the first six items but at this time point were restricted to 15 minutes to complete the problems presented as sums. The maximum achievable score was 18. At time point 3 a further nine items were included, two of which were edited so that they were presented in a format that the children would recognise. Children were again guided through the first six items and were then given 30 minutes to complete the problems presented as sums. The maximum achievable score was 27. A time limit was applied at the later time points to keep children engaged with the testing (the majority of children finished the task well within the time limit). See Appendix 1 for the full list of items.

4.2.4. Procedure

Children were tested in whole class groups at each time point, these ranged from 13 to 30 children. Other tasks were included at each time point but are not included here as they are not the focus of this chapter which investigates the predictive relationships of measures assessed at Time 1 on later untimed arithmetic achievement (for the full list of tasks presented to children at Time 1, 2 and 3 see Appendix 8, 11 and 12 respectively). For details on the procedure used at Time 1 see Chapter 2 (section 2.2.3, page 57). At the latter time points the WIAT arithmetic subtests were presented during the first of two testing sessions that lasted on average an hour each. Each child had their own booklet to mark their answers in.

4.3. Results

This study was interested in the predictive relationship of the Time 1 measures (age, nonverbal ability, vocabulary, comparison ability, number identification and arithmetic) on later arithmetic achievement. Children's results on the Time 1 predictor variables are presented in Chapter 2, therefore only the descriptive results for the Time 2 and 3 arithmetic outcome measures will be reported here. In exception to this further descriptive statistics are reported for the number identification task presented at Time 1 as in Chapter 2 it was found to be a significant predictor of children's arithmetic. The descriptive statistics are presented first followed by the exploration of the longitudinal predictors of children's arithmetic ability at Time 2 when they were in school Year 2 and aged 6 to 7 years old and then at Time 3 when children were in school Year 3 and aged 7 to 8 years old. Not all of the children completed each task due to attrition in the sample because of the longitudinal nature of the study and the multiple testing sessions required at each time point. The number of children that completed each of the measures at Time 1 is presented in Chapter 2 Table 2.4, at Time 2 163 children completed the WIAT Numerical Operations measure while at Time 2 162 children completed the task.

4.3.1. Descriptive Statistics

Children's performance on the WIAT Numerical Operations arithmetic task improved over time; Time 1 = 8.86 (*SD* = 2.47), Time 2 = 12.01 (*SD* = 2.79) and Time 3 = 15.57 (*SD* = 3.32). To investigate children's improvement on the measure statistically, a repeated-measures ANOVA was performed, with time (1-3) as the within subjects factor. This revealed that children's arithmetic scores improved significantly over time, $F(2, 304) = 554.14$, $p < .001$, $\eta_p^2 = .79$. Within-subject contrasts revealed that there was a significant increase in children's scores from Time 1 to Time 2, $F(1, 152) = 286.75$, $p < .001$, $\eta_p^2 = .65$ and from Time 1 to Time 3, $F(1, 152) = 932.97$, $p < .001$, $\eta_p^2 = .86$. Due to the significant growth in children's arithmetic skill from Time 1 to both Time 2 and Time 3 the possible predictors of this growth can be examined.

4.3.1.1. Number identification.

As number identification was found to be a significant concurrent predictor of children's arithmetic in Chapter 2, the errors that children made on the test were examined in more detail. Errors were classified into the following categories: syntactic, where children chose an option that included additional zeros e.g. 208 rather than 28; inversion, the option where digits were in the incorrect order e.g. 82 rather than 28; non responses and multiple answers and finally other, which represented errors such as choosing 23 rather than 235. It should be noted that 156 children completed the task and the values reported represent the number of children who chose that response. Table 4.3 shows that the majority of children were able to correctly identify one and two digit numbers and made the most errors when identifying the three digit numbers. In general where children made errors the most common was to choose the syntactic distractor (additional zeros) rather than the option where the digits were correct but in the incorrect order.

Table 4.3.
Detailed results for the number identification task

		Errors			
Target	Correct	Syntactic	Inversion	Non responses/ multiple answers	Other
6	155	n/a	n/a	1	-
14	137	n/a	16	3	-
28	139	12	2	3	-
52	132	13	5	6	-
76	142	5	5	2	2
235	34	116	1	3	2
163	26	122	6	2	-
427	28	123	2	3	-

4.3.2. Longitudinal Relationships

The longitudinal analyses with arithmetic at Time 2 and Time 3 will now be presented. First the correlations between the Time 1 measures and children's arithmetic achievement at Time 2 and 3 will be explored. This will be followed by longitudinal path models examining the predictors of this growth. In Chapter 2 a series of CFAs were performed to assess the factor structure of the comparison data, therefore rather than running correlation analyses with the raw scores from the nine comparison tasks, composite z-scores were computed based on the factor structure. Magnitude comparison judgement was a composite of the z-scores of the digit and nonsymbolic comparison tasks (digit far, digit close, NS far SS, NS close SS, NS 3:4 SA, NS 5:6 SA and NS 7:8 SA), and letter comparison a composite of the letter far and letter close tasks.

4.3.2.1. Correlations between Time 1 predictors and later WIAT arithmetic achievement.

The simple correlations are shown in Table 4.4. The relationships between the Time 1 measures and WIAT Numerical Operations at Time 2 are summarised first, and then with Time 3. All of the Time 1 measures were significantly and positively related to Time 2 arithmetic, although the strength of the correlations differed. The strongest relationship with Time 2 arithmetic was performance on the measure one year earlier but this was closely followed by magnitude comparison and number identification ability. There was also a moderate association with both letter comparison and nonverbal ability (Ravens). There were weak but significant relationships with age and vocabulary knowledge (BPVS). In general the strength of the relationships between Time 1 measures and arithmetic at Time 3 reduced. Children's Time 1 arithmetic skill was again the strongest correlate of their performance on the task at Time 3. There were moderate relationships with number identification, both of the comparison tasks and nonverbal ability (Ravens). The association with vocabulary knowledge was weak but significant and remained the same as with Time 2 arithmetic. Between these two time points there was no association of children's age and arithmetic skill. As children's arithmetic achievement at Time 1 was moderately related with achievement at both Time 2 and 3, this shows good stability and reliability of the measure in this group testing design but also the strength of the correlations meant that there was potential for other possible predictors to influence children's achievement on the task.

Table 4.4.
Correlations between Time 1 measures and later WIAT arithmetic achievement

Time 1 measures	Time 2	Time 3
Numerical Operations	.63**	.58**
Age	.24**	.09
Ravens	.41**	.39**
BPVS	.34**	.34**
Magnitude comparison	.57**	.49**
Letter comparison	.48**	.47**
Number identification	.57**	.53**

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

4.3.2.2. Longitudinal path models.

To investigate the causal relationships between the Time 1 measures and arithmetic at Time 2 and 3, the CFA model that was constructed from the Time 1 variables (age, nonverbal ability, vocabulary, magnitude comparison, letter comparison, number identification and the autoregressor WIAT Numerical Operations, shown in Chapter 2, Figure 2.11, page 72) was used as the measurement model to then construct longitudinal path models. All analyses were conducted using M*plus* and missing values were handled with Full Information Maximum Likelihood estimation.

4.3.2.2.1. Predicting Time 2 WIAT arithmetic achievement.

First, children's arithmetic achievement assessed one year later was predicted from the measures assessed at Time 1. Figure 4.1 shows the longitudinal path model, covariances between the Time 1 measures are not presented here for simplicity but are presented in Appendix 14. It should be noted that the values are almost identical to those in the Time 1 CFA (Chapter 2, Figure 2.11). The model provided a very good fit to the data, $\chi^2(128) = 171.57$, $p = .006$, *RMSEA* = .044 (90% *CI* = .025-.061), *CFI* = .966, *SRMR* = .046. As illustrated in the figure the only significant predictor of children's Time 2 WIAT arithmetic scores, over and above the autoregressor, was number identification ability. Altogether the predictors explained 82% of the variance in arithmetic skill at Time 2.

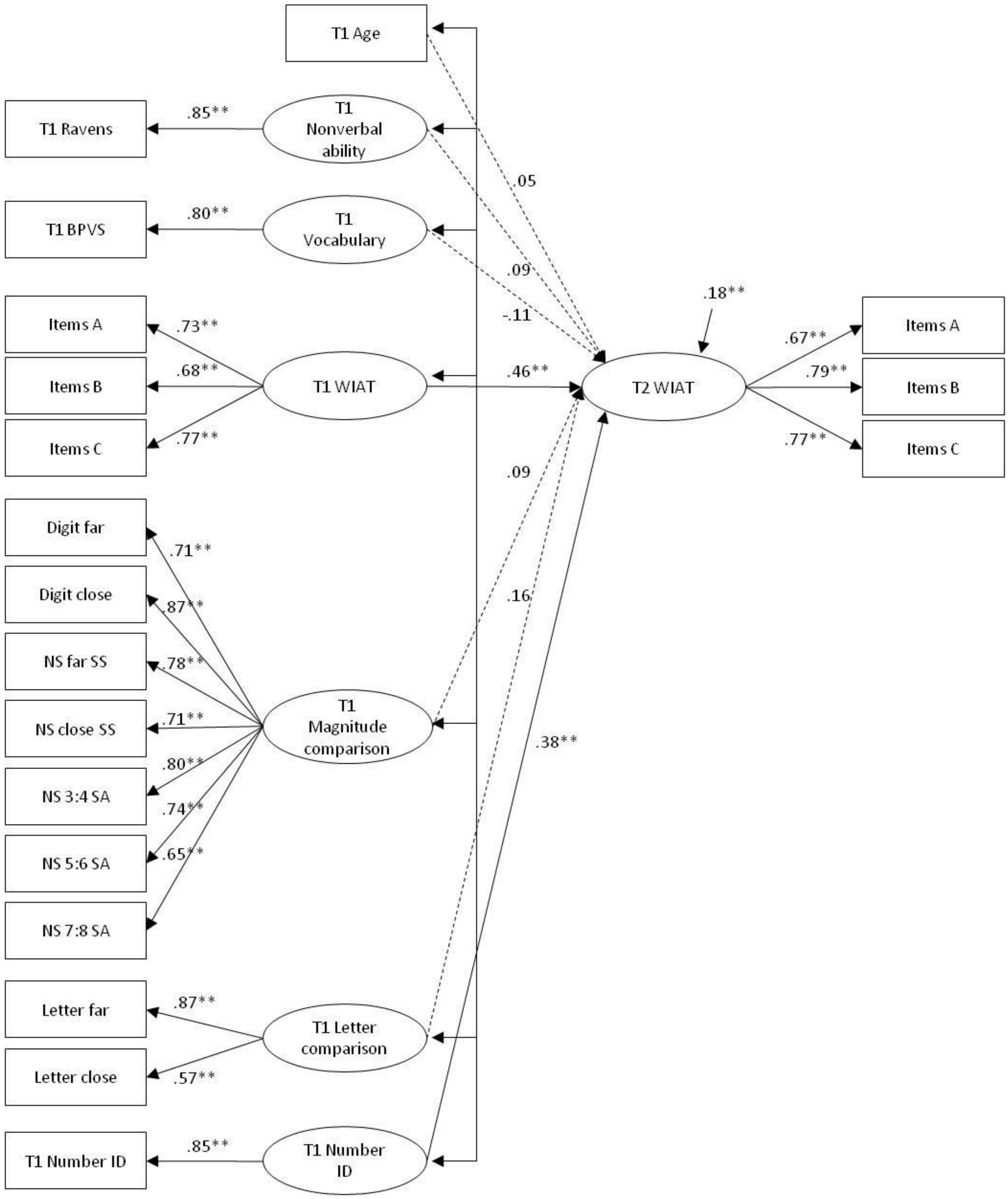


Figure 4.1. Path model predicting Time 2 WIAT Numerical Operations from Time 1 WIAT Numerical Operations, age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification. Model fit: χ^2 (128) = 171.573, *p* = .006, *RMSEA* = .044 (90% *CI* = .025-.061), *CFI* = .966, *SRMR* = .046.
* *p* < .05, ** *p* < .01

In previous studies, early arithmetic achievement has typically not been assessed and entered as an autoregressor, therefore the path analysis was also run without the autoregressor to see what pattern emerged. The CFA run in Chapter 2 (Figure 2.12 page 74) that included age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification was used as the measurement model to predict children's Time 2 WIAT arithmetic achievement (Numerical Operations). Figure 4.2 shows the longitudinal path model, covariances between the Time 1 measures are not presented here for clarity but are presented in Appendix 15. It should be noted that the values are almost identical to those in the CFA. The model provided an acceptable fit to the data, χ^2 (87) = 134.178, $p <$ .001, *RMSEA* = .056 (90% *CI* = .036-.074), *CFI* = .955, *SRMR* = .045. As in the previous longitudinal path model, the only significant predictor of children's Time 2 arithmetic was number identification at Time 1. Seventy three percent of the variance in Time 2 arithmetic skill was accounted for. It should be noted that when the autoregressor is removed from the analysis, the contribution of earlier magnitude comparison performance to later arithmetic achievement increases, although does not reach significance (p = .081). The contribution of number identification also increases (Beta weight (β) increases from .38 to .55).

4.3.2.2.2. Predicting Time 3 WIAT arithmetic achievement.

To investigate whether the same pattern of findings would hold over a two year time lapse (average = 21 months) children's arithmetic achievement at Time 3 was predicted from the measures at Time 1. Figure 4.3 shows the longitudinal path model which provided a very good fit to the data, χ^2 (128) = 162.834, p = .020, *RMSEA* = .040 (90% CI = .017-.057), *CFI* = .972, *SRMR* = .046 (covariances between the Time 1 measures are presented in Appendix 16 and were again almost identical to those in the Time 1 CFA shown in Chapter 2, Figure 2.11). As illustrated in the figure the only significant predictor of children's Time 3 WIAT Numerical Operations scores, over and above the autoregressor, was again number identification ability. Together the predictors explained 73% of the variance in arithmetic skill at Time 3. In summary whether predicting children's arithmetic achievement one year or two years later the pattern of predictors is very similar. This shows that children's number knowledge is important for not only their concurrent arithmetic achievement but also for their arithmetic development.

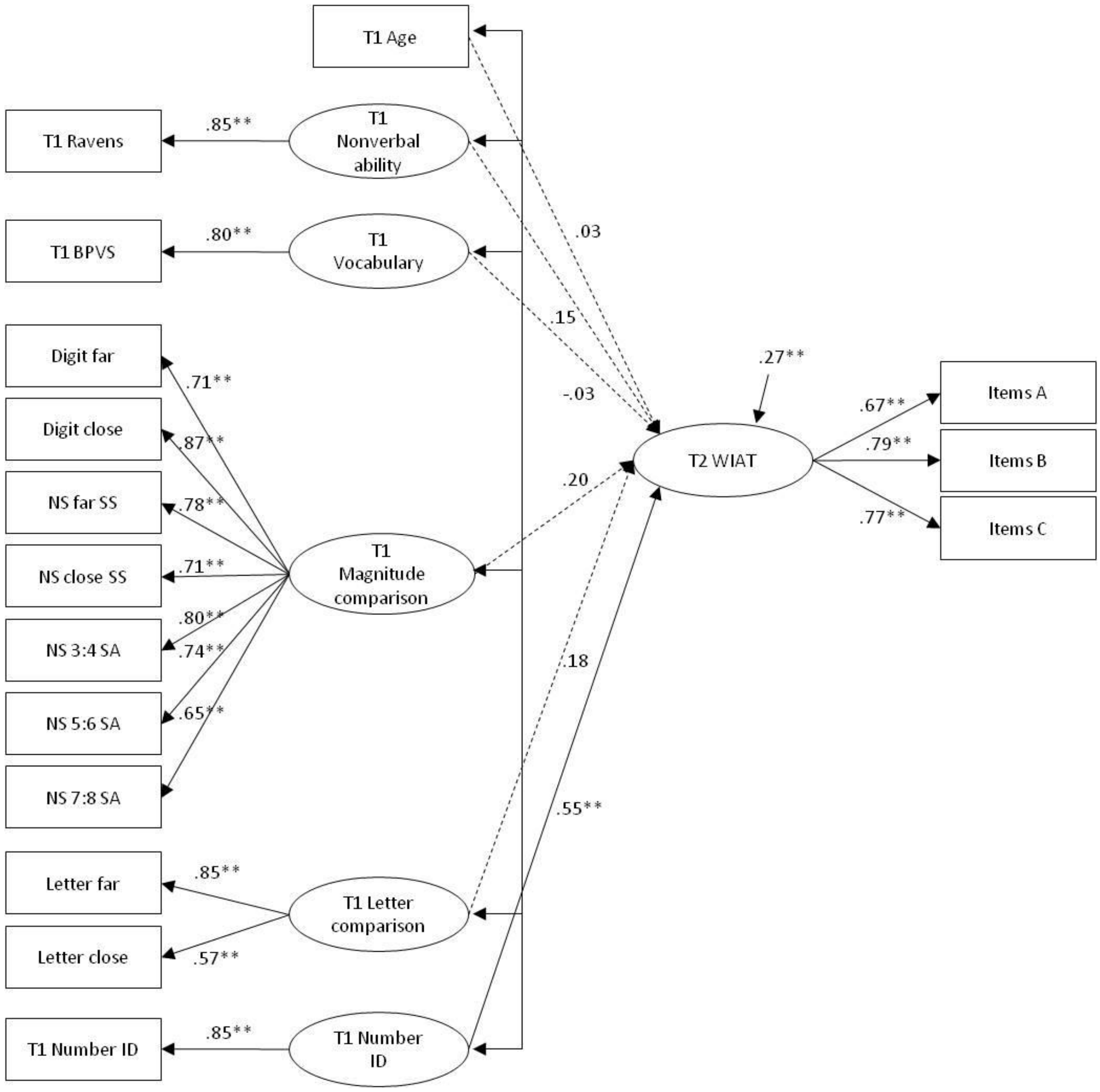


Figure 4.2. Path model predicting Time 2 WIAT Numerical Operations from Time 1 age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification skill (without autoregressor). Model fit: χ^2 (87) = 134.178, *p* < .001, *RMSEA* = .056 (90% *CI* = .036-.074), *CFI* = .955, *SRMR* = .045.
* *p* < .05, ** *p* < .01

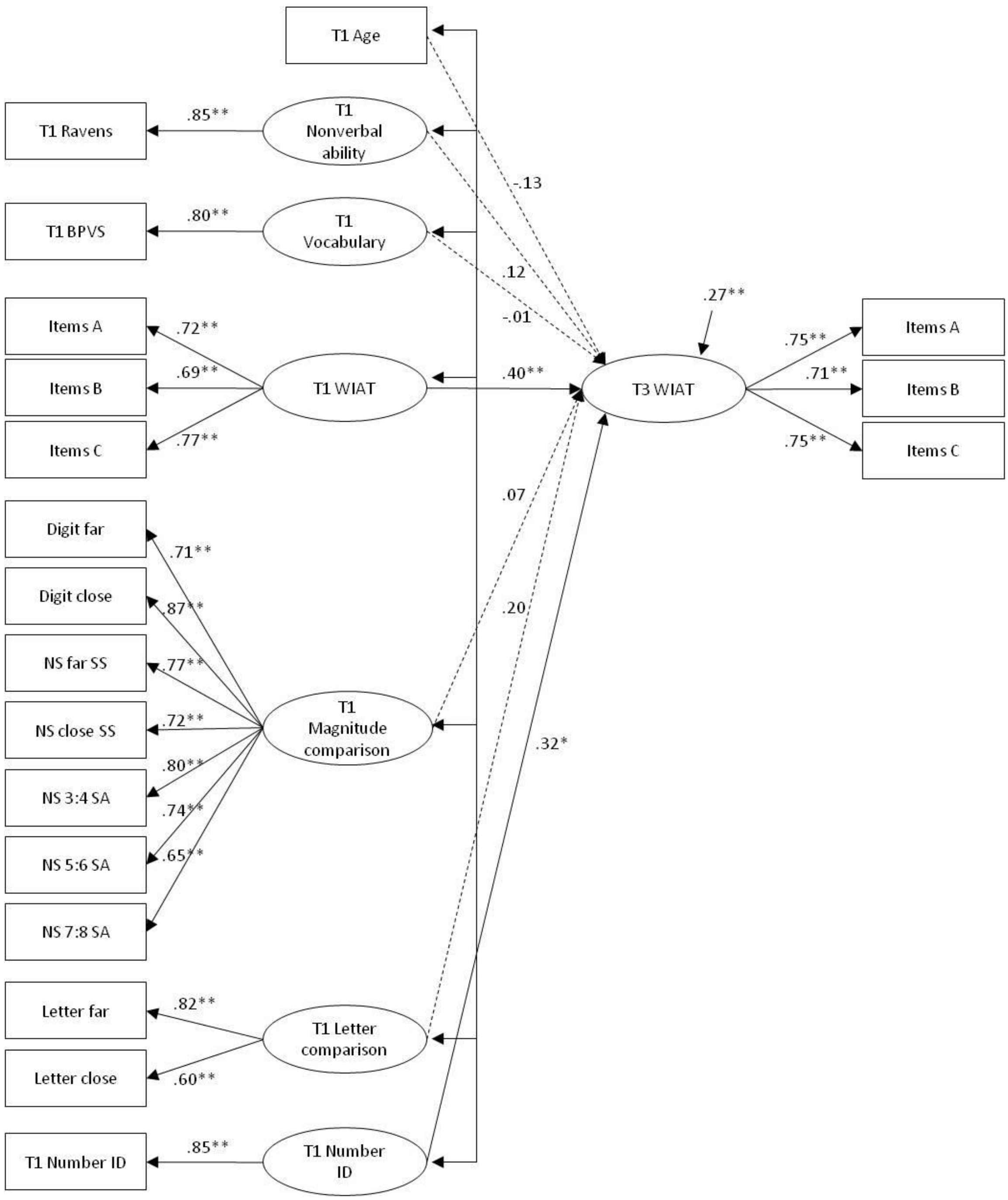


Figure 4.3. Path model predicting Time 3 WIAT Numerical Operations from Time 1 WIAT Numerical Operations, age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification. Model fit: χ^2 (128) = 162.834, p = .020, $RMSEA$ = .040 (90% CI = .017-.057), CFI = .972, $SRMR$ = .046.
* p < .05, ** p < .01

As with the Time 1 to Time 2 relationships the path analysis was run without the autoregressor to see if the same pattern emerged. The same CFA was again used as the measurement model. Figure 4.4 shows the longitudinal path model (covariances between the Time 1 measures are presented in Appendix 17, but it should be noted that the values are almost identical to those in the CFA). The model provided an acceptable fit to the data, χ^2 (87) = 124.256, *p* = .005, *RMSEA* = .050 (90% *CI* = .028-.069), *CFI* = .963, *SRMR* = .046. As in the previous longitudinal path models, the only significant predictor of children's Time 3 arithmetic was number identification at Time 1. The model explained 66% of the variance in children's Time 3 arithmetic achievement. It should be noted that when the autoregressor was removed from the analysis, children's age at Time 1 was a marginally significant predictor (*p* = .054), however the relationship was negative.

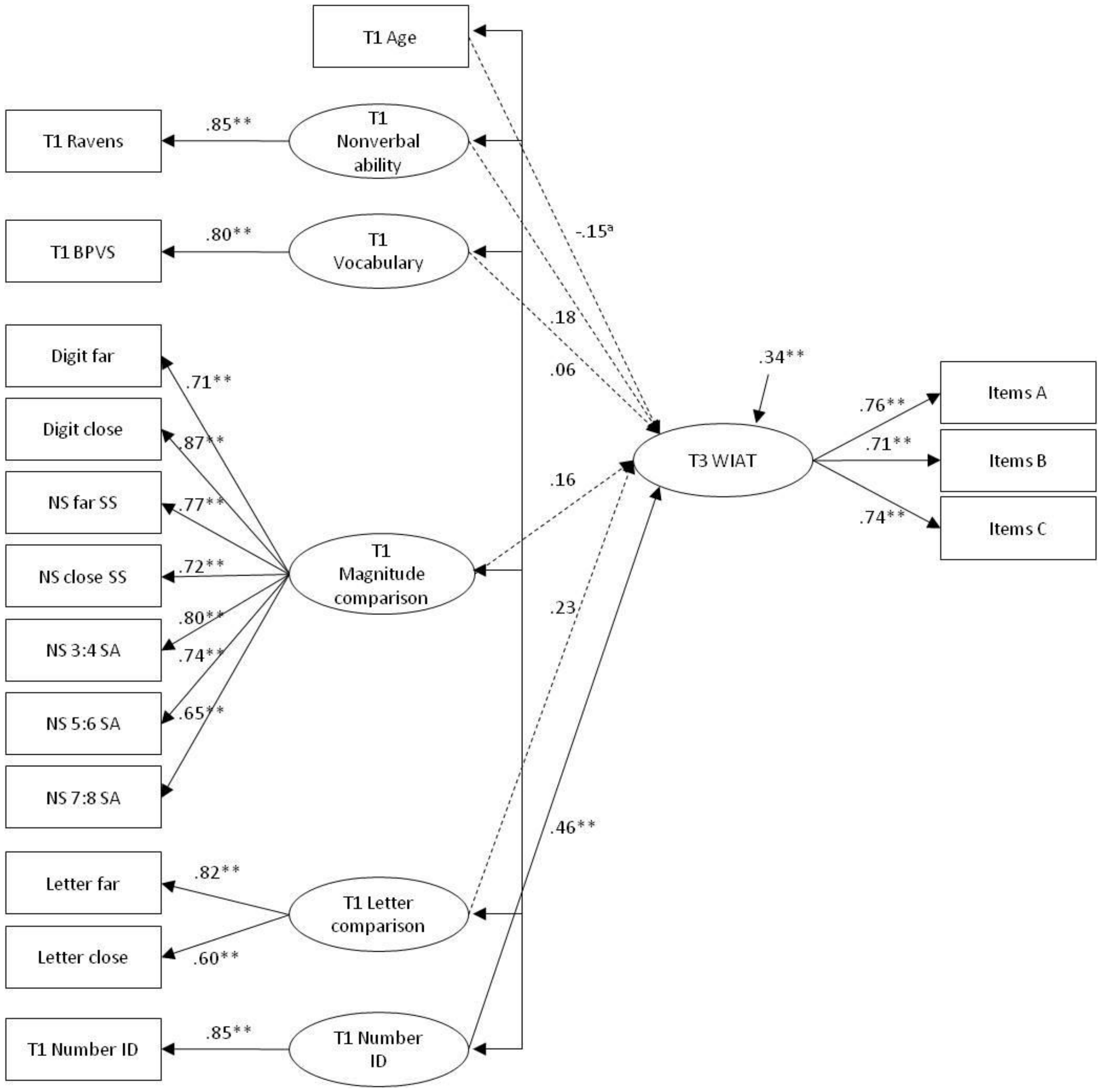


Figure 4.4. Path model predicting Time 3 WIAT Numerical Operations from Time 1 age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification skill (without autoregressor). Model fit: χ^2 (87) = 124.256, *p* = .005, *RMSEA* = .050 (90% *CI* = .028-.069), *CFI* = .963, *SRMR* = .046.

[a] *p* = .054, * *p* < .05, ** *p* < .01

4.4. Discussion

4.4.1. Summary of the Study

The aim of the current chapter was to investigate the longitudinal predictors of children's arithmetic achievement. Specifically, the question was whether children's magnitude comparison performance and ability to identify numbers would predict individual differences in later arithmetic skill, as was found concurrently at Time 1. Children involved in the previous study reported in Chapter 2 were retested roughly one and two years later on the WIAT Numerical Operations arithmetic measure which assesses basic number skills (e.g. identifying and writing numbers and counting) and basic calculation ability (e.g. addition, subtraction, multiplication and division problems). As there was little attrition in the sample the data were again analysed using sophisticated data analysis techniques (SEM).

4.4.2. Summary of the Results

The correlation analysis showed that both magnitude comparison and number identification ability were again associated with later arithmetic achievement. Arithmetic at Time 1 was also significantly correlated with later arithmetic. This indicated that the group presented measure of arithmetic has good stability but meant that there was potential for other tasks to predict development in arithmetic achievement.

A full path model was run to investigate which, if any, of the measures would predict significant variance in children's later arithmetic achievement. The CFA constructed in Chapter 2 which included possible predictor and control variables (age, verbal ability, nonverbal ability, magnitude and letter comparison, number identification, and earlier arithmetic ability) was used as the measurement model. It was found that the only significant predictor of children's arithmetic achievement at both Time 2 and 3 (along with the autoregressor) was their number identification ability at Time 1. This indicates that the findings are stable, with children's number identification skill at 5 and 6 years of age being important not only for their arithmetic achievement one year later but also almost two years later. In contrast, children's ability to compare magnitudes was not a significant predictor of later arithmetic achievement as it was concurrently.

4.4.3. Number Identification as a Predictor of Later Arithmetic Achievement

Children's performance on the number identification task was examined in more detail and it was found that when children made errors the most common one was to choose the symbol with additional zeros (syntactic error) rather than the option where the digits were correct but in the incorrect order, this is consistent with previous literature using number writing tasks (Power & Dal Martello, 1990; Seron & Fayol, 1994). In the written number system a child only has to learn 10 different symbols (0 to 9) but they must also learn that the position of a digit in the sequence determines its value (i.e. units, tens, hundreds etc) (Camos, 2008). The number identification task used here is most likely tapping children's knowledge of place value and knowledge of the rules of the system.

Children's ability to identify symbolic numbers predicts not only their current arithmetic achievement but was also found to predict individual differences in children's later arithmetic scores, this was true even when prior arithmetic skill was controlled. This shows that children's knowledge of the symbolic number system is important for arithmetic development from age 5 to 8 and that children who initially struggle with this kind of task may have more poorly developed arithmetic skills compared to their peers. This finding is in line with previous research with children with mathematical difficulties: children with lower mathematics ability perform more poorly on number reading tasks than children with mathematics ability in the normal range (De Smedt & Gilmore, 2011; Landerl et al., 2004; Mazzocco & Thompson, 2005). With regards to the literature on typically developing children, this finding is in contrast to some studies that only assessed children's ability to read smaller numbers (e.g. De Smedt et al., 2009; Soltész et al., 2010). However, it is in line with Gilmore et al. (2010) and Lembke and Foegen (2009) who found that children's Arabic number knowledge (assessed by naming one and two digit numbers) was significantly related to their mathematics and arithmetic ability.

The number identification task used in the current study could be more likened to a number transcoding task where children (and adults) are required to write numbers to dictation. The task could be seen as more difficult/demanding than simply reading single digit numbers; if a child was not able to directly retrieve the symbolic representation from long term memory, they may have needed to hold the verbal presentation in memory. Children also had to ignore the distractor items to choose the correct written symbol/code of the number. Children's ability to transcode or identify multidigit Arabic numerals also assesses their understanding of place-value. This finding that this number identification measure was a consistent predictor of arithmetic achievement provides support for Moeller et al. (2011) who found that in German speaking children the number of errors they made when transcoding one, two and three digit numbers in first grade was related to the number of errors they made on an addition task two years later. The findings of the present study suggest that tasks that involve children's knowledge of larger numbers (i.e. outside the one digit range) and assess place value understanding, are more useful measures to assess children's early arithmetic skill and for predicting later arithmetic achievement. It also extends the findings over a longer time period in English speaking children and shows that a measure of number knowledge predicts *growth* in arithmetic achievement.

The finding that knowledge of the Arabic number system is important for later arithmetic could be seen to be analogous to the importance of letter knowledge for reading development (e.g. Caravolas et al., 2012; Hulme, Bowyer-Crane, Carroll, Duff, & Snowling, 2012). In order for children to progress in their learning of these skills, in both domains they must first master the basic symbol set (i.e. Arabic numerals and letters) and their verbal labels. In the literacy domain it has been found that directly training children's letter-sound knowledge (alongside phoneme awareness) can improve reading skill (e.g. Hulme et al., 2012; Schneider, Roth & Ennemoser, 2000), therefore it may be advantageous to specifically train number knowledge in children who are experiencing difficulties with this and in numeracy in general. It should be noted that the process of identifying Arabic numbers is more complex that identifying letters. Multidigit numbers are combinations of the basic symbols 0 to 9 and therefore children need to understand place value and the rules of combining single digits in order to both read and write larger numbers.

4.4.4. Magnitude Comparison was Not a Significant Predictor of Later Arithmetic Achievement

The finding that magnitude comparison performance was a significant predictor of individual differences in children's arithmetic scores when they were 5 and 6 years old was not extended longitudinally. Although children's ability to compare magnitudes was found to be associated with their later arithmetic achievement it did not predict growth in their scores. This is in contrast to previous literature where both symbolic and nonsymbolic comparison ability has been found to predict later arithmetic ability (Desoete et al., 2012; De Smedt et al., 2009; Libertus et al., 2013; Mazzocco et al., 2011b; Reeve et al., 2012; however see Sasanguie et al., 2013; Sasanguie, Van den Bussche & Reynvoet, 2012). Possible reasons for the differences will be discussed in turn.

A major difference between the current study and those presented in Table 4.1 is the inclusion of the autoregressor. In the current study children's prior arithmetic attainment was included in the analyses so that it was development on the task that was being predicted. Sasanguie et al. (2013) found that children's efficiency at comparing numbers was a significant predictor of their later mathematics achievement, however when prior mathematics ability (in addition to age and spelling ability) was controlled symbolic comparison performance was no longer a significant predictor. In the present study the autoregressor was also removed from the path model to examine whether this was a reason for the conflicting results. Nonetheless, the same pattern emerged with number identification being the only significant predictor of children's later arithmetic achievement. Even with children's prior arithmetic skill not controlled, magnitude comparison did not predict later arithmetic achievement. A possible reason for the contrasting findings to Sasanguie et al. (2013) could be the inclusion of the number identification task in the current study. This again highlights the importance of assessing children's early knowledge of the symbolic number system.

The presentation of the comparison tasks differed between previous studies and this study which could have contributed to the conflicting findings. Typically comparison tasks are completed by children individually on a computer whereas in the current study the tasks were pencil and paper tasks completed in a group testing situation. However, in Chapters 2 and 3 it was found that the group presented comparison tasks replicated the effects of distance and ratio on performance that have been widely reported in the literature and magnitude comparison performance was also a concurrent predictor of children's arithmetic skill. In the current chapter, while magnitude comparison performance was not a predictor of later arithmetic a longitudinal association was observed. This therefore does not suggest that the comparison tasks themselves are leading to the inconsistent findings. The finding that magnitude comparison was a concurrent but not a longitudinal predictor of arithmetic achievement could be reflecting shared variance between the two measures at the first time point. It is also possible that at Time 1 there could be a stronger link between the skills involved, especially as children will have completed less items at Time 1 than at the later time points and these items are also less complex.

Due to the group presented design of the comparison tasks this meant that they measured both accuracy and speed with which comparisons were made (items correct within 30 seconds). Although both measures of performance have been found to be predictors of arithmetic it should be noted that typically performance on symbolic comparison tasks is measured by the effect of distance on comparison times (e.g. De Smedt et al., 2009; Reeve et al., 2012), while performance on nonsymbolic comparison tasks is a measure of children's accuracy at comparing items (e.g. prospective: Mazzocco et al., 2011b; retrospective: Halberda et al., 2008; Mazzocco et al., 2011a). De Smedt et al. (2009) found that while accuracy on the symbolic comparison task was related to later arithmetic achievement it was not a significant predictor of individual differences unlike the size of distance effect on comparison times (RTs). This point warrants further investigation and requires a study that investigates and analyses the relationship between both accuracy and speed performance on symbolic and nonsymbolic comparison tasks.

Linked to the previous point is the fact that the current study used a latent factor of symbolic and nonsymbolic comparison performance as a predictor, whereas the majority of studies have used one or the other. Very few studies have included both symbolic and nonsymbolic comparison tasks in longitudinal studies (see Desoete et al., 2012; Sasanguie, Van den Bussche & Reynvoet, 2012; Sasanguie et al., 2013). All of these studies have analysed the contribution of symbolic and nonsymbolic comparison separately and observed differences in the relationship with arithmetic between the two. Sasanguie and colleagues found that children's symbolic but not nonsymbolic comparison performance was a predictor of arithmetic fluency and mathematics one year later (Sasanguie et al., 2013; Sasanguie, Van den Bussche & Reynvoet, 2012). Whereas Desoete et al. (2012) found that the relationships changed over time; nonsymbolic comparison ability but not symbolic comparison ability predicted arithmetic achievement one year later, whereas the opposite pattern was observed when predicting arithmetic achievement two years later. In the current study the best fit to the comparison data at Time 1 was for symbolic and nonsymbolic comparison to form one factor, however it may still be important to analyse the importance of each separately in order to investigate whether a relationship with arithmetic and either one does exist. This will be explored in Chapter 6.

The statistical method of analysing the data varied between the current study and previous studies. Alongside the inclusion of the autoregressor, the current study used sophisticated data analysis techniques were used to analyse the data. Structural equation modelling techniques were applied to the data whereas typically regression techniques have been used. There are many benefits to this method, but one specific advantage is that it allows for the construction of latent variables, this estimates and removes random error variance making the analysis more valid and reliable (Werner & Schermelleh-Engel, 2009). Additionally, only a small number of studies control for children's general cognitive ability, therefore children's general ability could be reflected in the relationships between magnitude comparison performance and later arithmetic ability. However, it is acknowledge in some of the studies that don't control for general ability have found that it may not be important for arithmetic achievement (e.g. Desoete et al., 2012; Mazzocco et al., 2011b; Reeve et al., 2012).

The age at which the relationship between magnitude comparison performance and arithmetic achievement is assessed may lead to the differences in findings. First, the studies by Halberda et al. (2008) and Mazzocco et al. (2011a) are retrospective with magnitude comparison assessed at 14 years old and arithmetic achievement assessed earlier. Therefore in adolescents, superior magnitude comparison ability might be the consequence of better earlier arithmetic ability. Second, the children in the studies reported by Mazzocco et al. (2011b) and Libertus et al. (2013) were much younger than those involved in the present research. Early proficiency at comparing magnitudes may indeed be important for later arithmetic development but this relationship may be strongest when children are young (before formal schooling begins), whereas other skills (for example knowledge of symbolic numbers) may be more important when children have begun formal schooling. Future studies are needed to explore these proposals.

It is also feasible that the inclusion of the number identification task lead to the differences observed. Two of the studies included in Table 4.1 who reported that symbolic comparison ability was related to later arithmetic achievement also included a measure of children's symbolic number knowledge. De Smedt et al. (2009) found that neither accuracy nor speed of reading numbers was related to children's later arithmetic achievement. Reeve et al. (2012) also included a number naming task and found no difference on that measure between three groups of children who differed in their symbolic comparison speed (slow, medium and fast) and arithmetic scores. No correlation or regression analyses were reported in this study so we cannot be sure of the relationship between this measure and arithmetic achievement. The tasks used in these studies only assessed children's ability to read one digit numbers and the importance of assessing knowledge outside this range was discussed earlier, therefore more research is needed including both magnitude comparison tasks and number knowledge tasks.

4.4.5. Future Research Questions

The current study has found that children's early number knowledge is important for their arithmetic development. However, this study also found that children's ability to compare magnitudes (both symbolic and nonsymbolic) was not a predictor of later arithmetic achievement. Alongside calculation problems, the WIAT Numerical Operations arithmetic measure used here includes items that assess children's basic numerical skills such as identifying and writing numbers and counting. It is therefore possible that the predictive relationship of the number identification task reflects a strong relationship/similarity with these items. In order to be confident that this is not true it is important to also investigate its relationship with an arithmetic measure that assesses only calculation skill (Chapter 5).

It is also theoretically possible that different arithmetic operations depend on different precursor skills. With regards to solving arithmetic problems Dehaene and Cohen (1995; 1997) suggested that there are two possible routes; a direct asemantic route which would be used when answering over learned calculations, particularly multiplication facts, and a second indirect semantic route which would be used when solving subtraction and more complex addition problems. It is proposed that only the second route would involve magnitude representations in the approximate number system. As the WIAT Numerical Operations subtest included a range of operations and the items progressed from simple to more complex multidigit problems this question could not be addressed here. Chapter 5 will therefore investigate the relationship between magnitude comparison ability and the numerical operations (addition, subtraction and multiplication) separately.

A final possibility discussed earlier is that the relationship between magnitude comparison ability and arithmetic (and mathematics) may change over development. This question is beyond the scope of this thesis but it is important that future research addresses this. Linked with this is a question about the stability of the relationships between the comparison tasks (the CFA); it may differ early in development (i.e. preschool) and later ages (e.g. secondary school, adulthood). In order to clarify this, this question should be explored further using similar methods to those used in the current study.

Chapter 5.
Longitudinal Predictors of Arithmetic Fluency

5.1. Introduction

The current chapter explored the predictors of three of the numerical operations; addition, subtraction, and multiplication, separately. The three motives behind this are described in turn.

A longitudinal relationship between magnitude comparison and arithmetic (or mathematics) has previously been found in the literature but was not replicated in the study reported in Chapter 4, which assessed arithmetic using the WIAT Numerical Operations subtest. The differences between that study and those previously reported were discussed in Chapter 4; however, a further distinction could be the measure used to assess arithmetic achievement. Many different arithmetic and mathematics measures have been used by different authors; these range from speeded to untimed, some involve only calculations, whereas some include word problems. Therefore to fully understand the relationship between the possible predictor measures and later arithmetic achievement these arithmetic outcome measures need to be better defined. The current chapter examined the predictors of three numerical operations (addition, subtraction, and multiplication) separately, using measures that only contain calculation items in their symbolic form. The task was also time restricted with children being given one minute to complete as many items as they could, which would tap children's arithmetic fluency. This also allowed a contrast to the Numerical Operations arithmetic measure in which children were instructed to take their time with the items, although an overall time limit was applied at Time 2 and 3 (15 and 20 minutes respectively) in order to manage the pace of the testing session. It should be noted that the majority of children had completed the test within the time limit. Therefore, in contrast to the Numerical Operations test, the children were under a time pressure to complete problems on the one minute measures.

The WIAT Numerical Operations measure used to assess arithmetic in Chapter 4 includes items assessing basic numerical skills and calculation ability. It was found that number identification was a significant predictor of both Time 2 and Time 3 arithmetic achievement, whereas magnitude comparison performance was not. As the WIAT Numerical Operations measure consists of items which tap children's basic numerical skills (i.e. identifying numbers from a choice of numbers and letters, writing numbers, counting coins) it is possible that the predictive relationship of the number identification task reflects a strong relationship/similarity between these items. It is therefore important to use a measure that assesses only calculation skill. If the number identification measure remains as a significant predictor of children's later arithmetic ability then this task could be thought of as a measure of children's familiarity with Arabic symbols. Therefore arithmetic measures consisting of only calculation items were used in this study.

Third, it is theoretically possible that different arithmetic operations depend on different precursor skills. This could not be investigated with the WIAT, because it includes items from several arithmetic operations. This also links with the body of literature on patients with brain damage. There is evidence that some individuals with brain damage have selective impairments or preservations of arithmetic facts. Van Harskamp and Cipolotti (2001) reported on three patients each with a selective impairment on only one type of arithmetic fact problem. One patient had an impairment of retrieving multiplication facts, the second showed difficulty retrieving subtraction facts, and the third had an impairment in retrieving addition facts. Each had no difficulty on the other two numerical operations. Other studies have reported similar findings (e.g. multiplication impairment: Dehaene & Cohen, 1997; McCloskey, Aliminosa & Sokol, 1991; multiplication preservation: Delazer & Benke, 1997; subtraction impairment: Dehaene & Cohen, 1997; subtraction preservation: Dagenbach & McCloskey, 1992), although it should be noted that performance on the spared operation(s) are often not without error.

These studies suggest at least in adults that there are different routes to solving the different arithmetic operations otherwise all operations would be impaired. Dehaene and colleagues proposed the triple code model to explain numerical processing (e.g. Dehaene, 1992; Dehaene & Cohen, 1995; see literature review page 1 for a description of the triple code model) and according to this model there are two routes to solving arithmetic problems (see Dehaene & Cohen, 1997; for other views see McCloskey, 1992; Cipolotti & Butterworth, 1995). A direct asemantic route which would be used to answer over learned calculations, particularly multiplication facts. This involves the auditory verbal word frame as facts are stored as verbal associations in memory. The second is an indirect semantic route which would be used to solve subtraction and more complex addition problems. The operands are encoded as quantity representations which can then be manipulated in order to complete the problem. Though two separate routes are proposed the authors suggest that many calculations involve both routes. The triple-code model was proposed to explain adult numerical processing so caution should be taken when applying it to children's numerical processing. These studies also focus on the selective impairment or preservation of arithmetic facts rather than performance on complex calculations. Some children at 6 to 8 years of age will still need to calculate the answers to even simple sums (i.e. using the indirect route) as they do not have the answer stored in memory so are not able to use a retrieval strategy. These case studies are also on patients with damage to the brain who often experience other difficulties, for example with writing letters and words (Delazer & Benke, 1997) or comprehending spoken numbers (Dagenbach & McCloskey, 1992), so this may influence the impairments observed on calculations. This study is therefore an exploration into whether any differences in the importance of earlier skills exist between the different numerical operations.

With regards to the developmental literature, the importance of earlier skills for the different arithmetic operations is rarely investigated separately. With a focus on children's magnitude comparison ability and specifically the distance effect, Landerl and Kölle (2009) found that the effect of distance (as the regression slope) on symbolic comparison was significantly related to addition scores (r = .13) but not subtraction or multiplication scores (r = .11 and r = .03, respectively). Whereas the nonsymbolic distance effect (difference score) was significantly related to multiplication ability (r = -.14) while the associations with addition and subtraction ability only showed a trend towards significance (r = -.11 and r = -.10, respectively). It should be noted that all of the associations were weak in strength. These calculation tests were of a similar format to those presented in the current study (i.e. timed) although the items were more difficult (later items include two digit and three digit problems). Children were from grades 2 to 4 so ranged from 7 to 11 years old; although age was partialled out it is still possible that the relationship may differ over development with such a large age group.

Lonnemann and colleagues (2011) found that neither the symbolic nor the nonsymbolic distance effect was significantly related to children's addition (symbolic: r = -.27; nonsymbolic r = -.04) or subtraction (symbolic: r = -.12; nonsymbolic r = .01) ability. However, even though the relationship between symbolic comparison and addition did not reach significance it showed a stronger association than with subtraction (and compared to the relationships between nonsymbolic comparison and addition and subtraction). Again the sums were timed but more difficult than those used in the current study (included two digit problems) but children were of a more concentrated age range (8 to 10 years old).

In a study exploring magnitude comparison and individual differences in strategy use in third grade children (mean age = 8 years 10 months), Vanbinst et al. (2012) recorded the strategy used (retrieval, procedural or other) and reaction times for answering one digit addition and subtraction problems. It was found that the symbolic distance effect (slope) was significantly associated with the speed of retrieving arithmetic facts in both addition ($r = -.44$) and subtraction ($r = -.57$), and the speed of executing procedural strategies, again in both addition ($r = -.57$) and subtraction ($r = -.53$). This shows that children with smaller symbolic distance effects were faster at retrieving arithmetic facts and at using procedural strategies in both arithmetic operations. As no relationship was found between the nonsymbolic distance effect and arithmetic or mathematics achievement this was not investigated further. The focus of this study was on strategy use so it is different to those reported above and the current study where arithmetic was assessed with tests that gave children a set time limit to complete a given number of sums.

Although taken together the findings of the studies reported here are inconsistent, there is a suggestion that magnitude comparison ability may vary in its relationship with the different arithmetic operations, and therefore this warrants further investigation. All of the studies reported here investigated the relationship concurrently whereas it is important to investigate the longitudinal relationship between early comparison ability and later arithmetic achievement. The children involved in these studies were also older (ranging from 7 to 11 years old) than the children taking part in the current research (6 to 8 years old) so it will be also interesting to explore any differences in younger children who are at the early stages of their formal numeracy instruction.

Before hypotheses are made it is worth noting that linked with the ideas proposed by Dehaene and Cohen (1997) (the triple code model and the two routes for solving arithmetic problems) are the strategies that children may use to solve calculation problems. Dehaene and Cohen (1997) proposed that in adults over learned calculations would be solved via the direct asemantic route (i.e. retrieval), whereas subtraction and more complex addition problems would be solved via the indirect semantic route (i.e. procedural).

Addition is one of the first arithmetic operations to be taught in school (Department for Education and Skills, 2006) and a large body of research has focused on this. Addition can be seen as an extension of counting and children may initially solve addition problems by counting out all of the numbers in the sum (the sum model). They may at first use concrete examples of the addends, as in objects or their fingers, and may count aloud. Children then use a count on strategy, for example counting on from the first number in the problem and then move on to a more efficient count on strategy of identifying the larger of the two items and then counting on from that (the min strategy). With practice, children are able to retrieve answers stored in long term memory. It has been found that with age and experience children typically progress from using less to more efficient strategies (e.g. Farrington-Flint, Vanuxem-Cotterill & Stiller, 2009; Siegler, 1987), but there is also considerable heterogeneity and flexibility of strategy use not only between children of a similar age group but also within children (known as the overlapping waves theory; see Geary & Brown, 1991; Siegler, 1987; 1999). This is also true for the other operations (see Farrington-Flint et al., 2009; Lemaire & Siegler, 1995; Siegler, 1988, 1889). For example, Siegler (1987) reported that between kindergarten and second grade there was a decrease in the use of counting all and guessing strategies and an increase in the use of retrieval, but that within each age group at least five different strategies were being used. In a more recent study involving British children, Farrington-Flint et al. (2009) identified that the most commonly used strategy in both Year 1 and Year 2 children was to count on from the larger but that use of retrieval increased between Year 1 and Year 2 children. With regards to the other arithmetic operations a retrieval strategy has been found to be used less often to solve subtraction than addition problems (e.g. Barrouillet, Mignon & Thevenot, 2008) which would fit with Dehaene and Cohen’s (1997) suggestion that simple addition may be solved via a different route to subtraction. In multiplication Lemaire and Siegler (1995) observed that second grade children’s use of retrieval increased over a five month period (from 38 to 92% of problems). At the first testing point these children were on average 8 years 1 month old which is a similar age to the children involved in the current study at the third time point (8 years old). This would suggest that although Dehaene & Cohen’s model suggests that in adults multiplication problems are solved through a direct asemantic route this may not be true for younger children.

5.1.1. Hypotheses

In Chapter 4, magnitude comparison ability was not found to be a longitudinal predictor of children’s arithmetic achievement. It is possible that magnitude comparison ability may only predict certain types of arithmetic longitudinally. This study is therefore an exploration as there is little research examining the predictors of the different arithmetic operations in children. The adult patient literature will subsequently be used as a backdrop to make predictions about the patterns that might be seen in development. The current study will assess children’s addition, subtraction and multiplication ability using timed calculation tasks.

First, it is predicted that children's early arithmetic fluency skill (one minute addition) will be a predictor of their later arithmetic fluency achievement (addition, subtraction and multiplication).

According to the triple code model simple addition calculations would be solved by retrieving the answers from memory (using a fact retrieval strategy) and therefore using the direct asemantic route. Therefore we may not expect magnitude comparison ability to be related to addition achievement. However, due to the age of the children at Time 2 (6 and 7 years old) and Time 3 (7 and 8 years old) we may expect some children to be still using a counting strategy to answer the problems rather than using a fact retrieval strategy (see Farrington-Flint et al., 2009). The item difficulty also increases on the measure that will be used and the later items may go beyond what would be classed as 'over learned' calculations (i.e. they increase in problem size). As some children may be using a counting strategy the indirect semantic route would be engaged and therefore magnitude comparison ability may be important as children draw on their number representations to solve the problems. It is possible that a change in the importance of earlier skills might be seen between Time 2 and Time 3 addition as it would expected that later on more children have arithmetic facts stored in memory and are therefore using a retrieval strategy.

With regards to what skills are important for subtraction skill, according to the triple code model the indirect semantic route would be used to solve the calculations. In addition to this the retrieval strategy is used (or available) less often than with addition (e.g. Barrouillet et al., 2008). Children would manipulate quantity representations therefore it is hypothesised that magnitude comparison will be important for later subtraction skill.

In the adult literature it would be expected that multiplication problems could be solved by retrieving learnt arithmetic facts and therefore by the direct asemantic route. Due to children being taught multiplication after addition and subtraction the measure of multiplication skill was only presented at Time 3. It would be expected by then that a large number of children would be able to solve at least some of the calculations by retrieving the answers from memory and therefore the direct asemantic route would be used. For this reason magnitude comparison ability would be less important for later multiplication skill. However, some children might still need to work out the answer to the sum. This could involve running through the times tables but it is thought that these would still be stored as verbal associations and therefore not involve manipulating quantities. However, another way to solve the problem would be apply an addition strategy, for example 5 x 3 would involve adding 3 lots of 5 to get 15, this would then involve manipulating quantities. For these children it would then be expected that magnitude comparison ability would be important.

In Chapter 4, it was found that earlier number identification ability was the only significant predictor of later arithmetic ability, when prior arithmetic achievement was controlled for. If this really is an important skill for arithmetic development, and not just reflecting a strong association with some of the initial items on the Numerical Operations measure, then it is hoped that a relationship between number identification and each of the calculation measures is found.

In summary, Chapter 4 found that alongside earlier arithmetic skill, children's ability to identify Arabic numerals was important for their arithmetic development, while magnitude comparison performance, letter comparison performance, general ability and age were found not to be longitudinal predictors. The aim of this study was to explore the predictors of children's later calculation ability assessed by speeded arithmetic tests that focus on one numerical operation in order to examine whether any differences exist between them. The same CFA that was conducted in Chapter 2, page 78 was used as the measurement model to construct multiple longitudinal path models predicting children's addition, subtraction and multiplication ability separately.

5.2. Method

5.2.1. Design

There were three time points of assessment. Children were assessed at time point 1 between April and July 2010 (see Chapter 2 for concurrent data). Time point 2 took place between March and June 2011 which was on average 11.74 months later (range = 10.83 to 13.27 months). The third time point took place in February and March 2012 with an average time lapse of 9.63 months after Time 2 (range = 9.33 to 11.00 months) and 21.37 months after the first testing phase (range = 20.20 to 22.90 months).

5.2.2. Participants

Participants were the same children as those in Chapter 4 (see page 122) investigating the longitudinal predictors of the WIAT Numerical Operations arithmetic measure. At Time 1, 2 and 3 they were on average 6 years 3 months, 7 years 2 months and 8 years 0 months old, respectively.

5.2.3. Assessment Battery

At time points 2 and 3 children were retested on measures from the first time point and new tasks were also introduced. This chapter was interested in how children's performance on the measures administered at Time 1 predict their later ability on the different types of arithmetic operations assessed by speeded arithmetic tests (addition, subtraction and multiplication) therefore not all measures administered will be detailed here. The Time 1 variables that were used to predict later arithmetic achievement were nonverbal ability, vocabulary, number identification, comparison performance and speeded addition achievement, a detailed description of these measures can be found in Chapter 2 section 2.2.2 page 51. Regarding the Time 2 and 3 tasks only the speeded arithmetic tests are relevant for this study so they will be described in detail below.

5.2.3.1. One minute arithmetic tasks.

To assess speeded calculation skill (fluency) and children's ability at completing the different types of arithmetic operations, children completed addition, subtraction and multiplication tests. To assess addition and subtraction the same one minute tasks (Westwood et al., 1974) that were presented to children at Time 1 were presented again at Time 2 and Time 3. Children were given one minute to complete as many simple written addition problems (e.g., 2+1, 3+4) as possible (see Appendix 2) and one minute to complete as many written subtraction problems (e.g., 5-1, 6-2) as possible (see Appendix 3). Problems were presented as a list of 30 items; therefore a maximum score of 30 was possible for both tasks (items correct in 60 seconds). It should be noted that at Time 1 not all children completed the 1 minute subtraction task due to time constraints.

To assess children's multiplication ability, a one minute multiplication task was constructed for Time 3. Children begin to learn multiplication facts in Year 2 (2, 5 and 10 times-tables) and to use written methods to solve the multiplication problems (Department for Education and Skills, 2006). The multiplication task was therefore introduced at Time 3 as all children should have been able to attempt it. Problems from the 2, 5, and 10 times tables (e.g., 2x2, 6x5, 10x4) were selected. To be in line with the one minute addition and subtraction measures three problems from the 3 times-table (which children learn in Year 3) were added to the end of the test to increase the number of items to 30. The sums were presented as a list of 30 items in the same format as the other speeded arithmetic measures (see Appendix 18 for the list of items). A maximum score of 30 was possible (scored as items correct in 60 seconds). On all measures at all Time points any correct response where an individual numeral was reversed was scored as correct (i.e. 3 as Ɛ, or 15 as 1Ƨ).

5.2.4. Procedure

For details on the procedure used at Time 1 see Chapter 2 (section 2.2.3, page 57). Children were tested in whole class groups at all time points. The speeded arithmetic tasks completed at Time 2 and Time 3 were presented alongside other measures which are not detailed here (for the full presentation order at Time 2 see Appendix 11 and for Time 3 see Appendix 12). At Time 2 the size of each class ranged from 13 to 29 children and at Time 3 ranged from 15 to 30 children. At the latter time points the arithmetic measures were presented during the first of two sessions that lasted on average an hour each. At all time points each child had their own booklet to mark their answers in.

5.3. Results

This study explored the predictive relationships of the Time 1 measures (age, nonverbal ability, vocabulary, comparison ability, number identification) on later arithmetic achievement both with and without controlling for children's prior arithmetic ability. Children's results on the predictor variables are presented in Chapter 2 (page 57), therefore only the results for the Time 2 and 3 outcome measures will be reported here. The descriptive statistics are presented first followed by the exploration of the longitudinal predictors of children's arithmetic ability at Time 2 when they were in school Year 2 and aged 6 to 7 years old, and then at Time 3 when children were in school Year 3 and aged 7 to 8 years old. Children were tested as whole class groups over three sessions at Time 1 and two sessions at Times 2 and 3 therefore not all children were present for all measures. Due to this being a longitudinal study there was some attrition in the sample. The number of children who completed each arithmetic task is reported for each measure.

5.3.1. Descriptive Statistics

Children's performance on each of the timed arithmetic tasks is presented in Table 5.1. Performance on each test is presented as raw scores and these represent the number of items correct within one minute. Table 5.1 shows that there was a wide range of scores on all of the calculation tasks and children's performance improved with development.

5.3.1.1. Frequency of errors.

Due to the one minute arithmetic tests requiring the children to complete the problems under a time pressure it was decided to explore the number of errors children made to check whether this speeded element induced large numbers of errors or not. Table 5.2 displays the number of errors that children made on each one minute test at each time point. In general the mean number of errors was low on all tasks, however, while some children made no errors at all, some children made large numbers of errors, particularly on the one minute subtraction and multiplication measures. To explore this further the frequency of errors is displayed in Table 5.3. It can be seen that on each one minute test the majority of children made no errors. On every task apart from one minute addition at Time 2, at least one child made more than four errors. However, the number of children who did make larger numbers of errors was quite low (less than 10 percent of children).

Table 5.1.
Descriptive statistics on the Time 1, 2 and 3 speeded arithmetic measure

	Time 1					Time 2					Time 3				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
1 minute addition	165	7.80	4.28	0 (4)	23	162	12.47	5.17	2	26	162	16.38	5.86	1 (1)	30 (1)
1 minute subtraction	79	6.86	3.82	0 (6)	15	163	8.57	3.99	0 (4)	19	161	11.73	5.00	0 (3)	29
1 minute multiplication	-	-	-	-	-	-	-	-	-	-	162	12.80	6.78	0 (1)	30 (1)

Note. Figures in parentheses represent the number of children scoring at floor and ceiling.

Table 5.2.
Number of errors made on the Time 1, 2 and 3 *speeded arithmetic measures*

	Time 1					Time 2					Time 3				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
1 minute addition	165	0.87	1.30	0	8	162	0.40	0.70	0	3	162	0.49	1.02	0	9
1 minute subtraction	79	1.39	1.96	0	9	163	1.06	1.88	0	13	161	0.73	1.41	0	10
1 minute multiplication	-	-	-	-	-	-	-	-	-	-	162	0.93	1.71	0	12

Table 5.3.

Frequency of errors on the Time 1, 2 and 3 speeded arithmetic measures (percentages)

	Time 1		Time 2		Time 3		
Number of errors	1 minute addition	1 minute subtraction[a]	1 minute addition	1 minute subtraction	1 minute addition	1 minute subtraction	1 minute multiplication
0	52.73	44.30	70.37	48.47	69.14	58.39	55.56
1	27.27	21.52	20.99	30.06	20.37	27.95	25.93
2	8.48	17.72	6.79	11.66	8.02	8.70	8.64
3	7.88	6.33	1.85	4.29	0.62	1.24	3.70
4 or more	3.64	10.13	0.00	5.52	1.85	3.73	6.17

Note. [a] Not all children completed this measure at Time 1, N = 79

5.3.2. Improvement of Arithmetic Skill

From Table 5.1 it can be seen that children's scores on the addition and subtraction tests increased over time. To investigate children's improvement on the addition measure statistically, a repeated-measures ANOVA was performed, with time (1-3) as the within subjects factor. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(2) = 35.64$, $p < .001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .82$). There was a significant effect of time, $F(1.65, 245.46) = 338.89$, $p < .001$, $\eta_p^2 = .70$, indicating that children's scores did in fact increase with age. Within-subject contrasts revealed that there was a significant increase in children's scores from Time 1 to Time 2, $F(1, 149) = 282.85$, $p < .001$, $\eta_p^2 = .66$ and from Time 1 to Time 3, $F(1, 149) = 463.88$, $p < .001$, $\eta_p^2 = .76$.

To investigate children's improvement on the subtraction measure a repeated-measures ANOVA was performed, with time (1-3) as the within subjects factor. Not all children completed the subtraction task at Time 1 therefore this analysis is only performed on a subset of children's data (n = 74). Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(2) = 11.09$, $p = .004$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .88$). Again, a significant effect was observed, $F(1.75, 127.76) = 89.41$, $p < .001$, $\eta_p^2 = .55$, with children's scores increasing over time. Within-subject contrasts revealed that the increase in children's scores from Time 1 to Time 2 was a significant, $F(1, 73) = 44.23$, $p < .001$, $\eta_p^2 = .38$, as was the increase from Time 1 to Time 3, $F(1, 73) = 129.42$, $p < .001$, $\eta_p^2 = .64$.

5.3.3. Longitudinal Relationships

The longitudinal analyses with arithmetic at Time 2 and Time 3 will now be presented. First the correlations between the Time 1 measures that were identified as possible predictors of later arithmetic achievement and the Time 2 (addition and subtraction) and Time 3 (addition, subtraction and multiplication) outcome measures will be explored. This will then be followed by longitudinal path models examining the predictors of later arithmetic fluency. In Chapter 4 the predictive relationships of the measures assessed at Time 1 were examined both with and without controlling for the autoregressive effect so this will be repeated here with the one minute arithmetic measures.

5.3.3.1. Correlations between Time 1 predictors and later calculation achievement.

When looking at the relationship of the comparison tasks with later arithmetic achievement the correlations were carried out with composite scores of the comparison tasks rather than the raw scores. This is because in Chapter 2 a series of CFAs were performed to assess the factor structure of the comparison data and revealed a 2 factor structure was the best fit to the data. Composite z-scores were computed based on this final factor structure; magnitude comparison judgment was a composite of the digit and nonsymbolic tasks (digit far, digit close, NS far SS, NS close SS, NS 3:4 SA, NS 5:6 SA and NS 7:8 SA), while letter comparison was a composite of letter far and letter close.

The simple correlations are shown in Table 5.4 Children's performance on both of the Time 1 arithmetic measures was moderate to strongly related with their later achievement on the same task. This shows good stability and reliability of the measures in this group testing design, while the strength of the correlations meant that there was potential for other tasks to predict growth in children's achievement. The strength of the associations of the two Time 1 tests (addition and subtraction) with the later arithmetic measures (addition, subtraction and multiplication) are also similar.

At Time 2, excluding the associations with earlier performance on the arithmetic measures, the strongest association with children's addition scores was earlier magnitude comparison performance but this was closely followed by their number identification ability. There were also moderate correlations with letter comparison and nonverbal ability, while the association with vocabulary scores was weak. There was no relationship with age. With one minute subtraction the strongest correlates were magnitude comparison and letter comparison performance. The correlations with number identification and Ravens were moderate in strength. There were weak, but significant, relationships with age and BPVS scores.

The pattern of relationships is similar for all three categories of arithmetic measures at Time 3. There were no significant associations between children's age at Time 1 and any of the Time 3 arithmetic tasks. Regarding children's scores on the addition the task, their earlier performance on the number identification measure was the strongest correlate but there were also moderate associations with magnitude comparison. The relationship with letter comparison is weak to moderate, while the association with nonverbal ability (Ravens) became weak in strength, as was the relationship with children's vocabulary knowledge. Regarding children's subtraction ability there were moderate correlations with earlier magnitude comparison, letter comparison and number identification scores, and the relationship with nonverbal ability could also be considered moderate in strength. The association with vocabulary was again weak in strength. The relationship between children's multiplication scores and early vocabulary knowledge was slightly stronger than with addition and subtraction (although would still only be considered weak to moderate in strength). The relationships with earlier magnitude comparison, letter comparison and number identification scores were all moderate in strength and the association with nonverbal ability was weak.

Table 5.4.
Correlations between Time 1 measures and Time 2 and Time 3 arithmetic

	Time 2		Time 3		
Time 1 measures	Addition	Subtraction	Addition	Subtraction	Multiplication
Age	.13	.15*	.09	.09	.14
Ravens	.38**	.40**	.30**	.38**	.34**
BPVS	.25**	.25**	.26**	.20*	36**
Magnitude Comparison	.59**	.53**	.48**	.50**	.45**
Letter comparison	.45**	.50**	.36**	.46**	.40**
Number identification	.53**	.44**	.55**	.50**	.51**
1 minute addition	.75**	.61**	.57**	.60**	.53**
1 minute subtraction[a]	.65**	.66**	.62**	.65**	.61**

Note. [a] Max n = 76
* $p < .05$, ** $p < .01$

5.3.3.2. Longitudinal path models (controlling for earlier calculation ability).

To explore whether the longitudinal predictors of addition, subtraction and multiplication differ, path models were constructed for each measure separately. All SEM analyses were conducted using M*plus* and missing values were handled with Full Information Maximum Likelihood estimation. The CFA constructed from the Time 1 variables age, nonverbal ability, vocabulary, magnitude comparison, letter comparison, number identification and one minute addition (shown in Chapter 2, Figure 2.14, page 78) was used as the measurement model. To enable comparisons with the exploration of children's development on the WIAT arithmetic measure, children's prior arithmetic skill needed to be controlled for. As not all children completed the one minute subtraction task at Time 1 and the one minute multiplication test was only introduced at Time 3, children's performance on the one minute addition task was used as the control for prior arithmetic skill in all analyses. The correlations between early addition and later subtraction and multiplication scores were also moderate in strength (see Table 5.4) which shows that there was an association between the different measures.

To simplify the presentation of the models all covariance values between the Time 1 measures have been removed from the figures but are acknowledged by the multiple headed arrow that connects to each Time 1 factor. Unless otherwise stated the values are almost identical to those in the Time 1 CFA presented in Chapter 2 (Figure 2.14). The covariance values are presented in the Appendix and each is referred to in the corresponding section. The predictors of Time 2 addition and subtraction skill will be explored first, followed by the predictors of Time 3 addition, subtraction and then multiplication skill.

5.3.3.2.1. Addition achievement at Time 2.

First, the possible contributors to children's growth in addition skill were explored. Figure 5.1 shows the longitudinal path model, covariances between the Time 1 measures are presented in Appendix 19. The model provided an acceptable fit to the data, $\chi^2(68)$ = 113.468, p < .001, *RMSEA* = .062 (90% *CI* = .041-.082), *CFI* = .956, *SRMR* = .042. As illustrated in the figure there were four significant predictors of children's Time 2 speeded addition scores, these were the autoregressor, number identification, magnitude comparison and age. Whilst the relationships between Time 2 addition and Time 1 addition, number identification and magnitude comparison were all positive, the relationship with age was negative. It is interesting that children's magnitude comparison performance is a significant predictor of their later addition fluency but did not predict unique variance in scores on the untimed arithmetic outcome measure that assessed children's basic numerical abilities and included other numerical operations (as in Chapter 4). Together these predictors explained 68% of the variance in addition skill at Time 2.

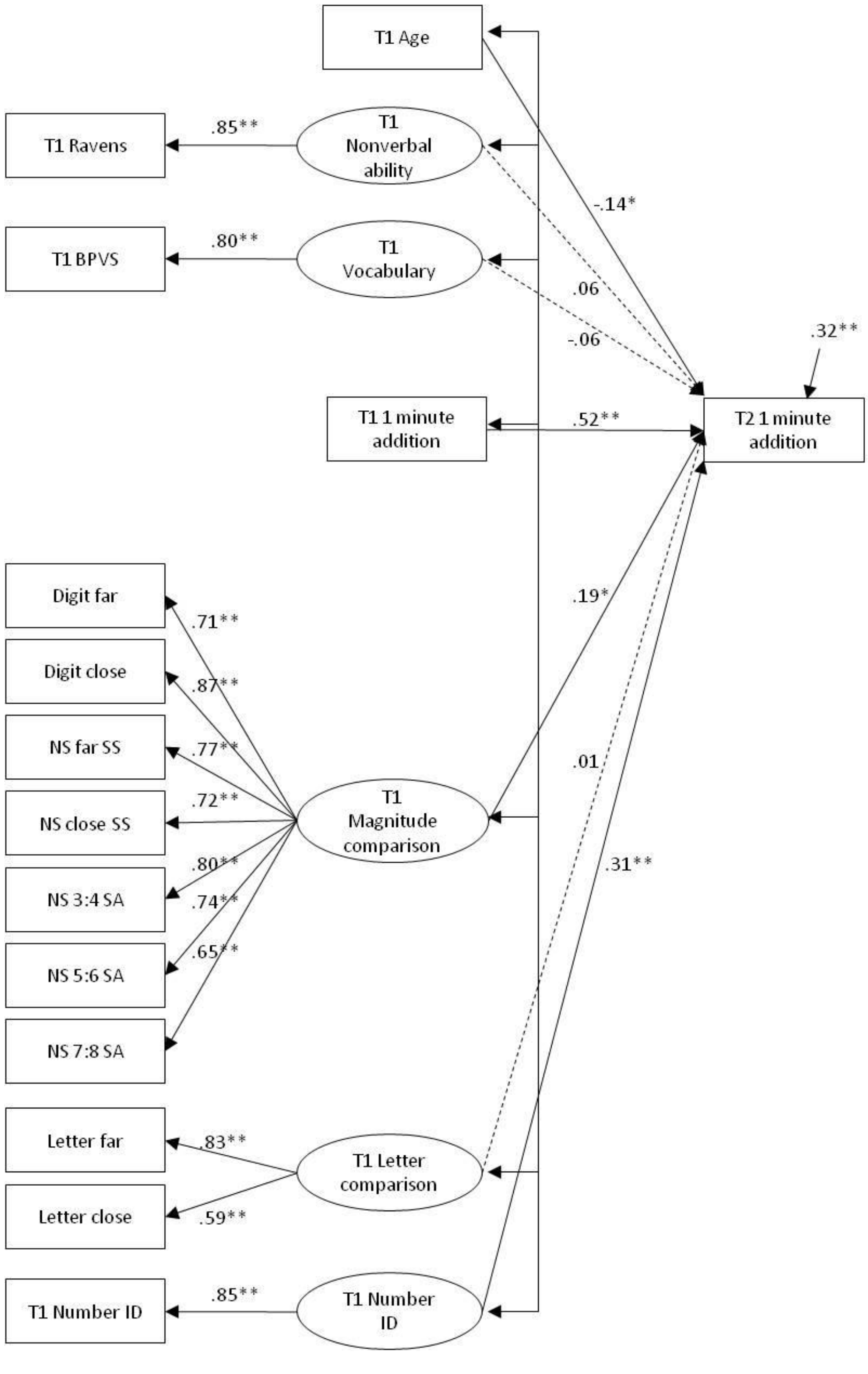


Figure 5.1. Path model predicting Time 2 one minute addition from Time 1 one minute addition, age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification. Model fit: χ^2 (68) = 113.486, p < .001, *RMSEA* = .062 (90% *CI* = .041-.082), *CFI* = .956, *SRMR* = .042.
* p < .05, ** p < .01

5.3.3.2.2. Subtraction achievement at Time 2.

Time 2 subtraction achievement was then predicted by the Time 1 variables. Due to over half the sample not being administered the one minute subtraction measure at Time 1 one minute addition achievement was used to control for earlier calculation ability. Figure 5.2 shows the longitudinal path model, covariances between the Time 1 measures are presented in Appendix 19. The model provided an acceptable fit to the data, χ^2 (68) = 122.162, p < .001, *RMSEA* = .068 (90% *CI* = .048-.087), *CFI* = .945, *SRMR* = .043 and explained 52% of the variance in subtraction achievement. However, the only significant predictor of children's subtraction fluency was their achievement on the speeded addition measure one year earlier. This is very different to addition achievement which was predicted by number identification, magnitude comparison and age, alongside prior addition ability.

5.3.3.2.3. Addition achievement at Time 3.

To investigate whether the same pattern of findings would hold over a longer time lapse children's addition achievement assessed at Time 3 (when children were 7 to 8 years old) was predicted from the measures assessed at Time 1. The model shown in Figure 5.3 provided an acceptable fit to the data, χ^2 (68) = 111.768, p < .001, *RMSEA* = .061 (90% CI = .040-.081), *CFI* = .955, *SRMR* = .042 and predicted 54% of variance in children's addition scores (covariances between the Time 1 measures are presented in Appendix 20). Alongside the autoregressor there was only one significant predictor of addition achievement at Time 3, this was number identification. The Beta weight (β) of number ID, and therefore the predictive power has increased from Time 2 to Time 3 (Beta weights .31 and .52 respectively). Children's age and magnitude comparison ability are now no longer significant predictors of later addition achievement when the time lapse is increased to almost two years. However, the contribution of magnitude comparison ability to later addition skill is marginal (p = .68) and it should be noted that the beta weights remained quite stable (predicting Time 2: β = .19, predicting Time 3: β = .20).

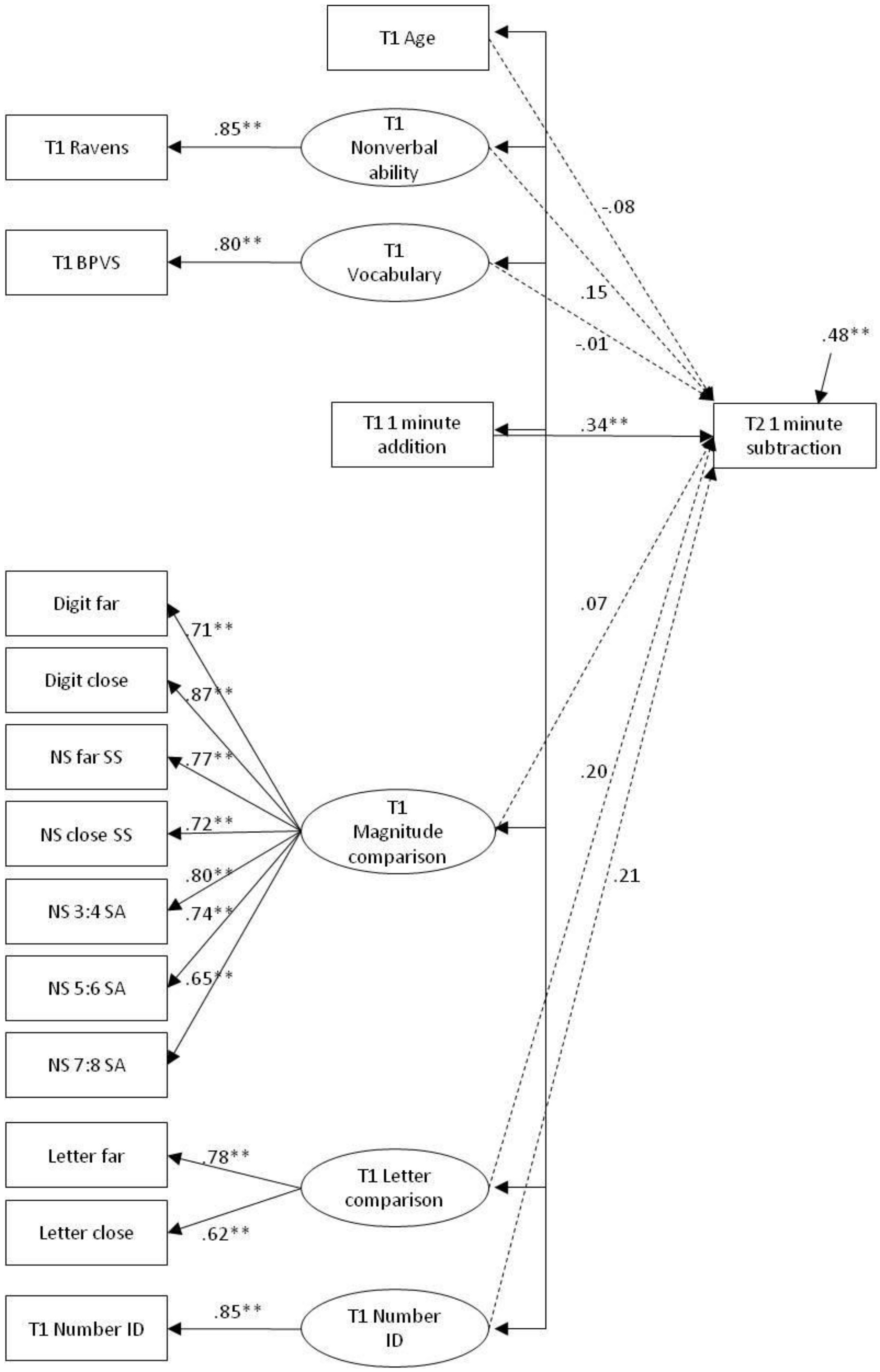


Figure 5.2. Path model predicting Time 2 one minute subtraction from Time 1 one minute addition, age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification. Model fit: χ^2 (68) = 122.162, *p* < .001, *RMSEA* = .068 (90% *CI* = .048-.087), *CFI* = .945, *SRMR* = .043.
* p < .05, ** p < .01

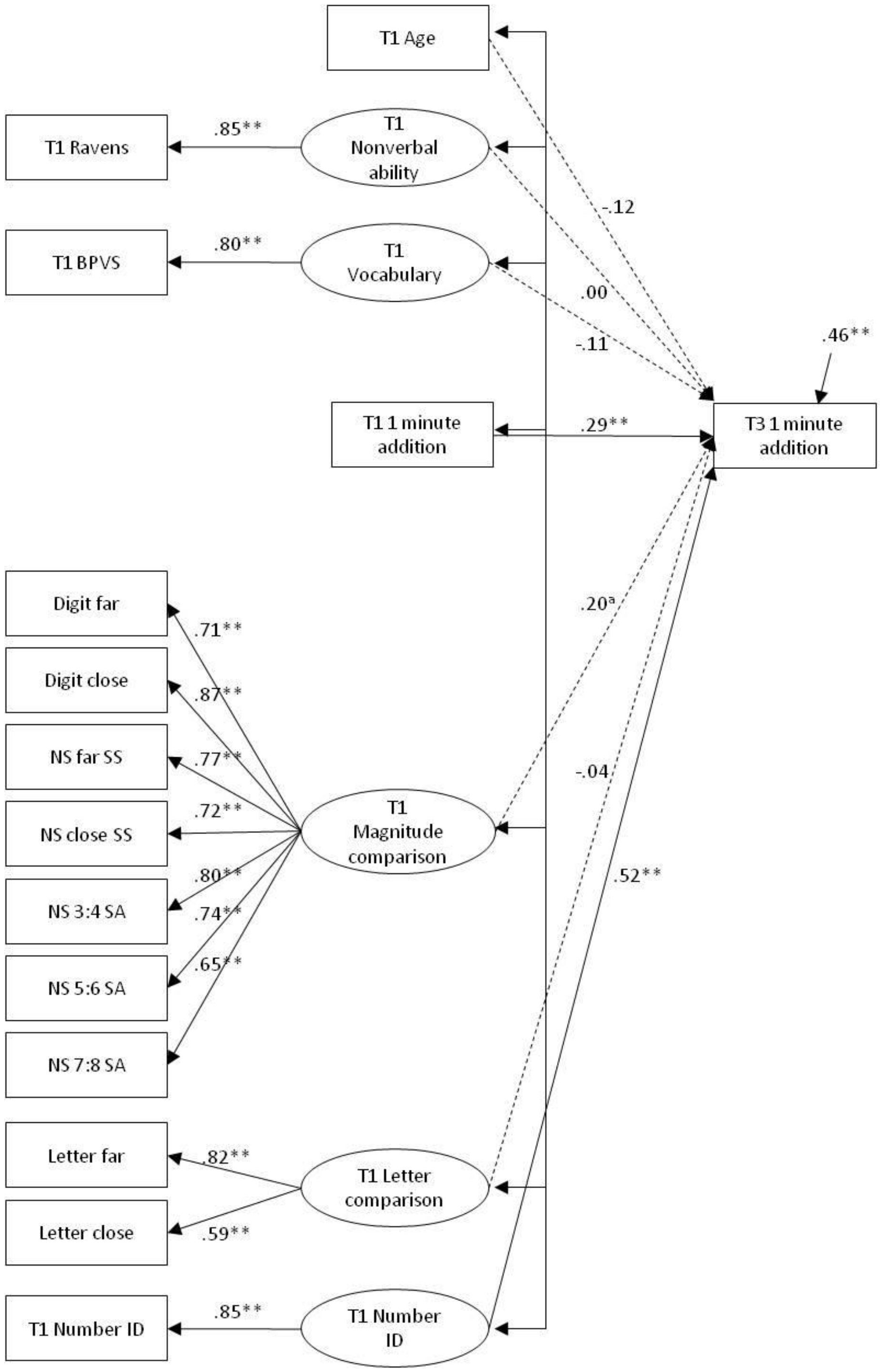


Figure 5.3. Path model predicting Time 3 one minute addition from Time 1 one minute addition, age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification. Model fit: χ^2 (68) = 111.768, $p < .001$, *RMSEA* = .061 (90% CI = .040-.081), *CFI* = .955, *SRMR* = .042.

[a] $p = .068$, * $p < .05$, ** $p < .01$

5.3.3.2.4. Subtraction achievement at Time 3.

The analysis was repeated predicting children's subtraction achievement from the Time 1 measures included in the CFA. The model provided a reasonable fit to the data, χ^2 (68) = 123.298, *p* < .001, *RMSEA* = .069 (90% CI = .049-.088), *CFI* = .944, *SRMR* = .044 and accounted for 52% of variance in children's subtraction scores. Covariances between the Time 1 measures are presented in Appendix 20. As can be seen in Figure 5.4, children's earlier achievement on the addition task was again a significant predictor of variance in children's later subtraction ability. In contrast to children's subtraction achievement at Time 2 there was an additional significant predictor, this was their number identification ability. This finding is consistent with those found when predicting children's addition achievement only and when arithmetic achievement is assessed by a standardised untimed measure which includes multiple numerical operations (WIAT).

5.3.3.2.5. Multiplication achievement at Time 3.

At time point 3 children completed a timed multiplication test so at this time point children's multiplication fluency could be predicted from the Time 1 measures. Akin to predicting individual differences in later addition and subtraction scores, children's performance on the one minute addition task was used to control for earlier calculation ability. The model (depicted in Figure 5.5) provided an acceptable fit to the data, χ^2 (68) = 112.472, *p* < .001, *RMSEA* = .061 (90% CI = .040-.081), *CFI* = .954, *SRMR* = .043. Covariances between the Time 1 measures are presented in Appendix 20. The findings replicated those when predicting children's addition and subtraction scores at Time 3, with the only significant predictors of later multiplication ability being children's early addition skill and number identification ability. Forty eight percent of variance in children's multiplication scores was explained.

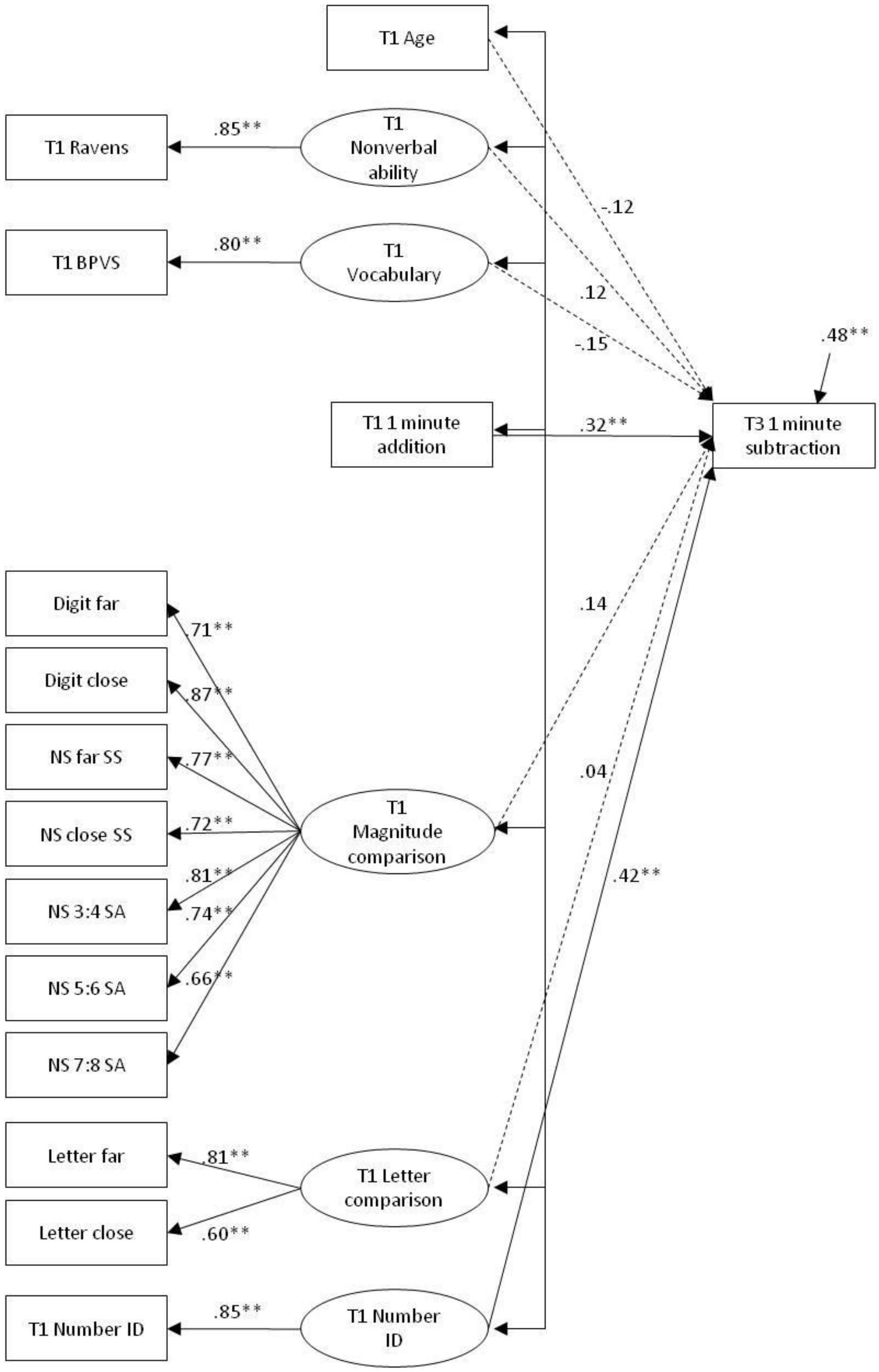


Figure 5.4. Path model predicting Time 3 one minute subtraction from Time 1 one minute addition, age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification. Model fit: χ^2 (68) = 123.298, *p* < .001, *RMSEA* = .069 (90% CI = .049-.088), *CFI* = .944, *SRMR* = .044.
* *p* < .05, ** *p* < .01

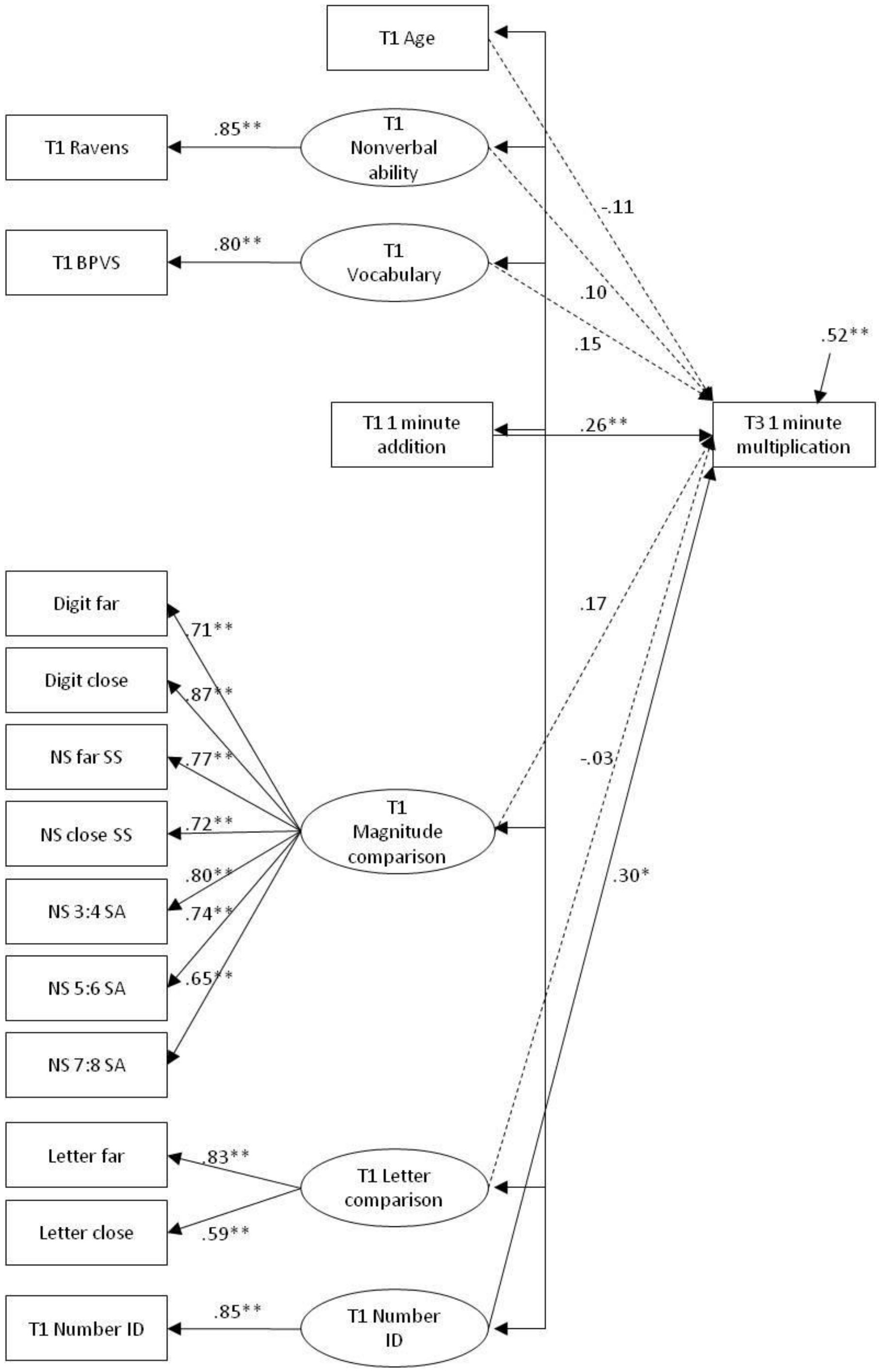


Figure 5.5. Path model at predicting Time 3 one minute multiplication from Time 1 one minute addition, age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification. Model fit: χ^2 (68) = 112.472, p < .001, *RMSEA* = .061 (90% CI = .040-.081), *CFI* = .954, *SRMR* = .043.
* p < .05, ** p < .01

5.3.3.2.6. Summary: Predicting later calculation ability when controlling for prior calculation ability.

As expected children's early arithmetic ability, measured using a timed addition test, predicted children's later ability to solve addition, subtraction and multiplication problems. Children's number identification ability was also important for later addition, subtraction and multiplication ability, except subtraction ability at Time 2. Interestingly, magnitude comparison performance predicted individual differences in children's addition scores one year later and still showed a trend towards significance two years later. However, it was not a predictor of later subtraction or multiplication ability. Children's nonverbal ability, vocabulary knowledge and ability to compare letters were not predictive of their later addition, subtraction or multiplication scores.

5.3.3.3. Longitudinal path models (without controlling for prior calculation ability).

In Chapter 4 the importance of the suggested predictors of later arithmetic ability were also explored without controlling for children's prior arithmetic achievement. This was to enable comparisons of the results found in this research with those in the existing literature as children's early arithmetic knowledge is generally not controlled for. In the current chapter this analysis was repeated with the three different numerical operations. The CFA that was run in Chapter 2 (Figure 2.12, page 74) that included only age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification was used as the measurement model for these analyses. In keeping with the presentation of the analyses including the control, the predictors of Time 2 addition and subtraction ability will be explored first, followed by the predictors of Time 3 addition, subtraction and multiplication skill.

5.3.3.3.1. Addition achievement at Time 2.

Children's achievement on the one minute addition test at Time 2 was predicted by the Time 1 measures age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification ability. The model provided an acceptable fit to the data, $\chi^2(61) = 92.396$, $p = .006$, $RMSEA = .055$ (90% CI = .030-.076), $CFI = .965$, $SRMR = .041$ and is illustrated in Figure 5.6 (covariances presented in Appendix 21). When children's earlier performance on the addition test was not controlled for there were only two significant predictors of later addition skill; magnitude comparison and number identification. This shows that whether or not children's prior addition skill is controlled for, both the ability to compare magnitudes (symbolic and nonsymbolic) and knowledge of symbolic numbers are important for later ability to complete addition calculations. Fifty three percent of the variance in Time 2 addition scores was explained. There was one difference to the when the autoregressor was included, age was no longer a significant predictor.

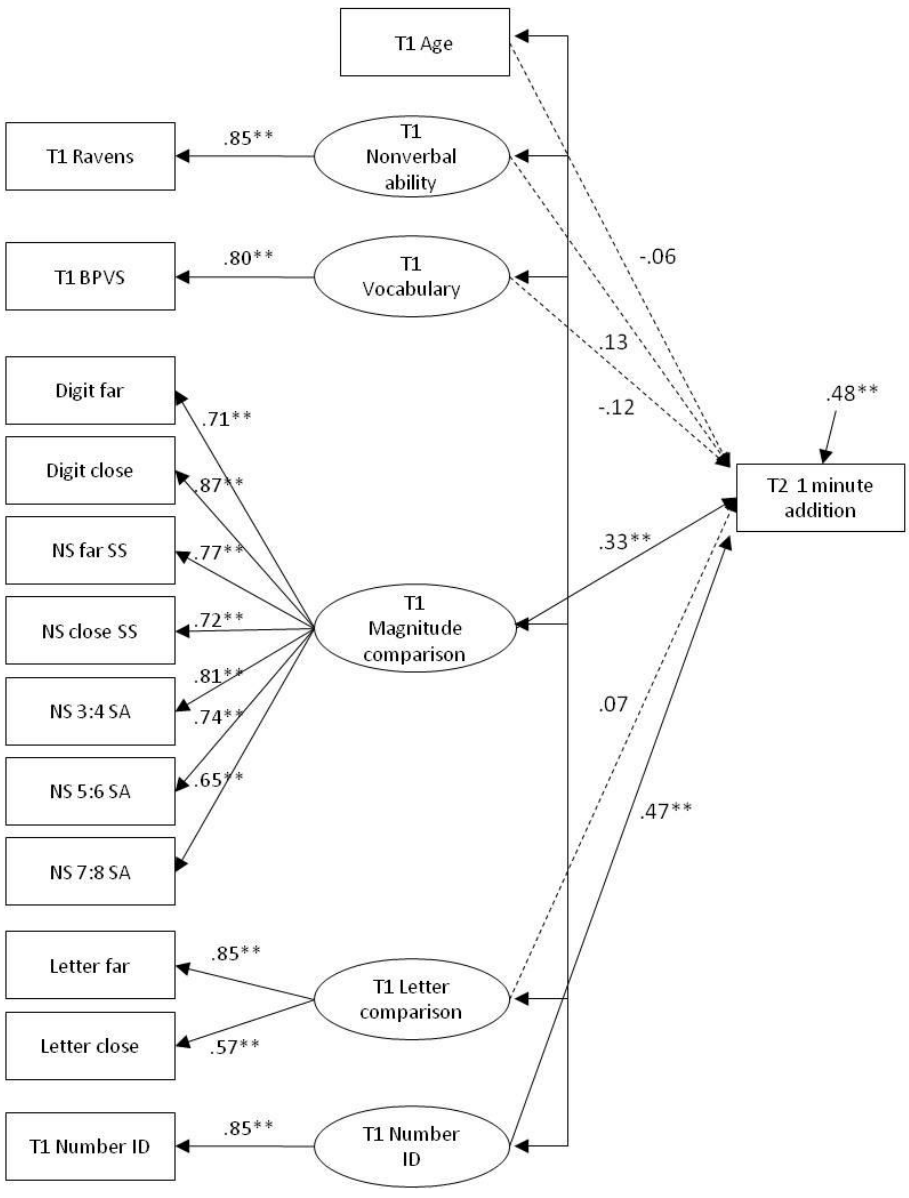


Figure 5.6. Path model predicting Time 2 one minute addition from Time 1 age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification. Model fit: χ^2 (61) = 92.396, *p* = .006, *RMSEA* = .055 (90% *CI* = .030-.076), *CFI* = .965, *SRMR* = .041.
* *p* < .05, ** *p* < .01

5.3.3.3.2. Subtraction achievement at Time 2.

The predictors of children's achievement on the one minute subtraction test at Time 2 were then explored. The model provided an acceptable fit to the data, χ^2 (61) = 103.443, *p* < .001, *RMSEA* = .063 (90% *CI* = .042-.084), *CFI* = .952, *SRMR* = .042 and is illustrated in Figure 5.7 (covariances presented in Appendix 21). When children's addition ability was controlled for in the analysis it was the only significant predictor of children's subtraction achievement one year later. When this was removed children's number identification became the only significant predictor of later subtraction ability, with children who had better number identification ability at Time 1 gaining higher subtraction scores one year later. This model explained 46% of the variance in Time 2 subtraction scores.

5.3.3.3.3. Addition achievement at Time 3.

To investigate whether the same pattern of findings would hold over a longer time lapse children's addition achievement assessed at Time 3, when children were 7 to 8 years old, was predicted from the measures assessed at Time 1. Figure 5.8 illustrates the path model, which provided a good fit to the data, χ^2 (61) = 89.877, *p* = .010, *RMSEA* = .052 (90% *CI* = .027-.074), *CFI* = .967, *SRMR* = .041. Covariances are presented in Appendix 22. Reflecting the findings when predicting children's Time 2 addition achievement, there were two significant predictors of Time 3 addition ability: magnitude comparison and number identification, which explained 49% of the variance.

5.3.3.3.4. Subtraction achievement at Time 3.

The predictors of children's Time 3 subtraction ability were then explored. The model (see Figure 5.9) provided an acceptable fit to the data, χ^2 (61) = 107.626, *p* < .001, *RMSEA* = .066 (90% *CI* = .045-.087), *CFI* = .947, *SRMR* = .044 (covariances are presented in Appendix 22). Children's number identification ability was again the only significant predictor of children's subtraction ability; this is the same pattern as with Time 2 subtraction skill. There was a trend towards magnitude comparison performance being a significant predictor (*p* = .069). The model explained 47% of the variance in Time 3 subtraction scores.

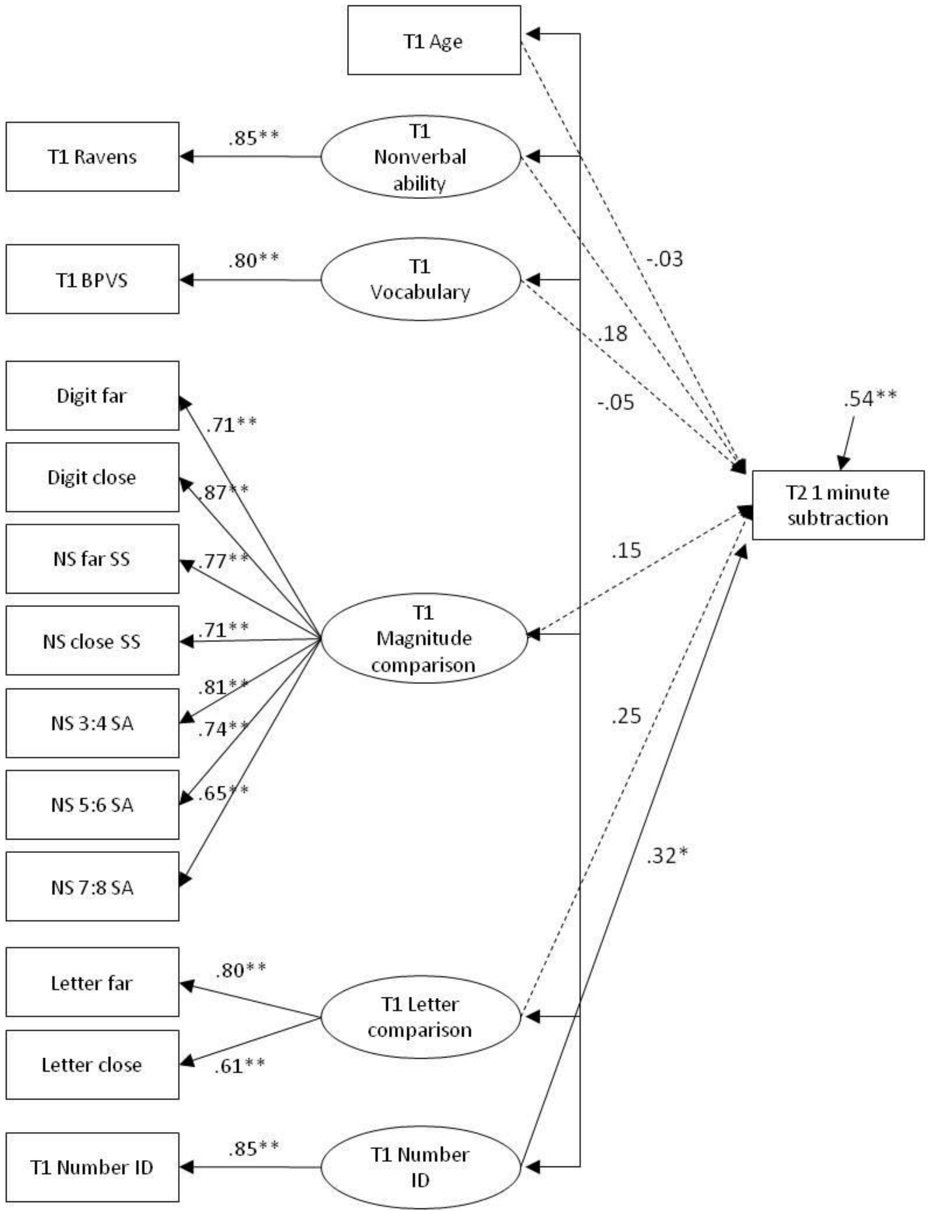


Figure 5.7. Path model predicting Time 2 one minute subtraction from Time 1 age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification. Model fit: χ^2 (61) = 103.443, p < .001, *RMSEA* = .063 (90% *CI* = .042-.084), *CFI* = .952, *SRMR* = .042.
* p < .05, ** p < .01

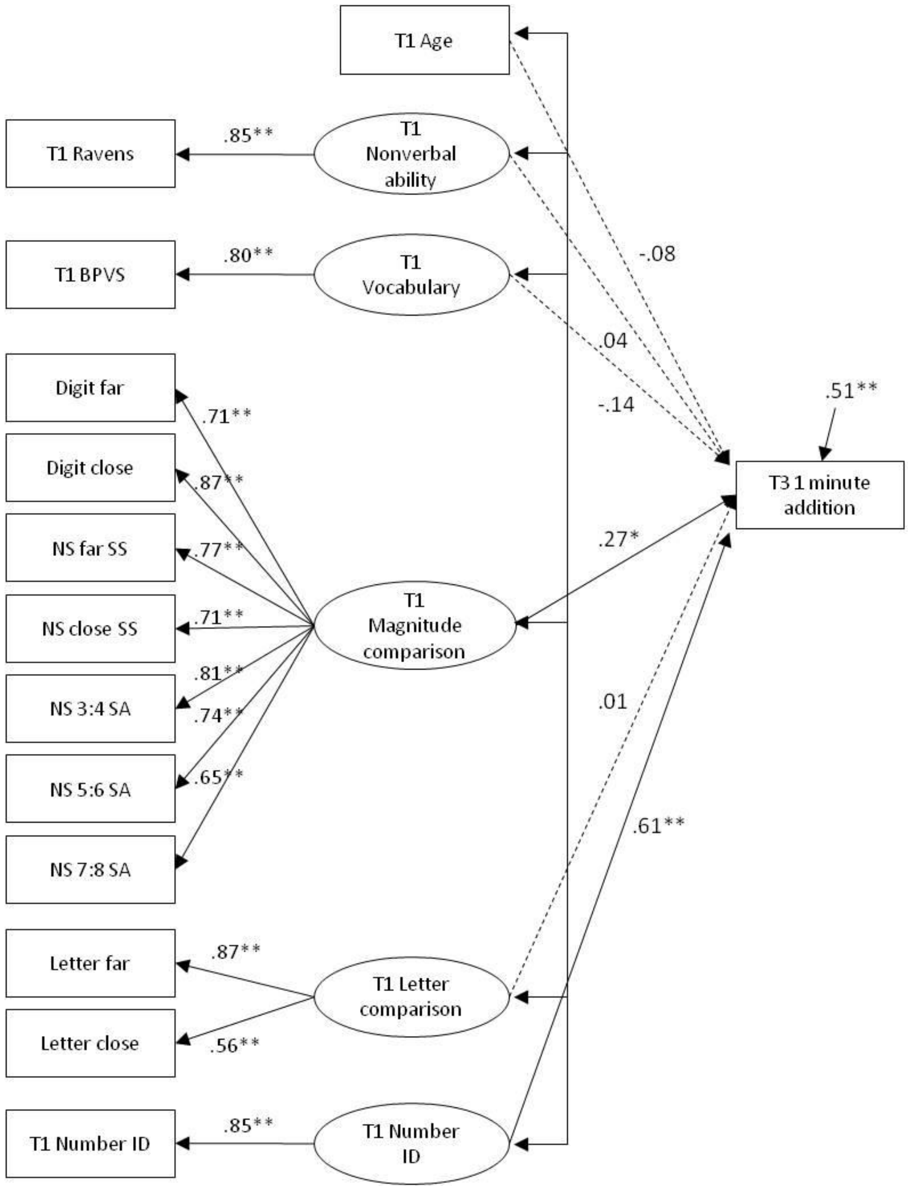


Figure 5.8. Path model predicting Time 3 one minute addition from Time 1 age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification. Model fit: χ^2 (61) = 89.877, *p* = .010, *RMSEA* = .052 (90% *CI* = .027-.074), *CFI* = .967, *SRMR* = .041.

* *p* < .05, ** *p* < .01

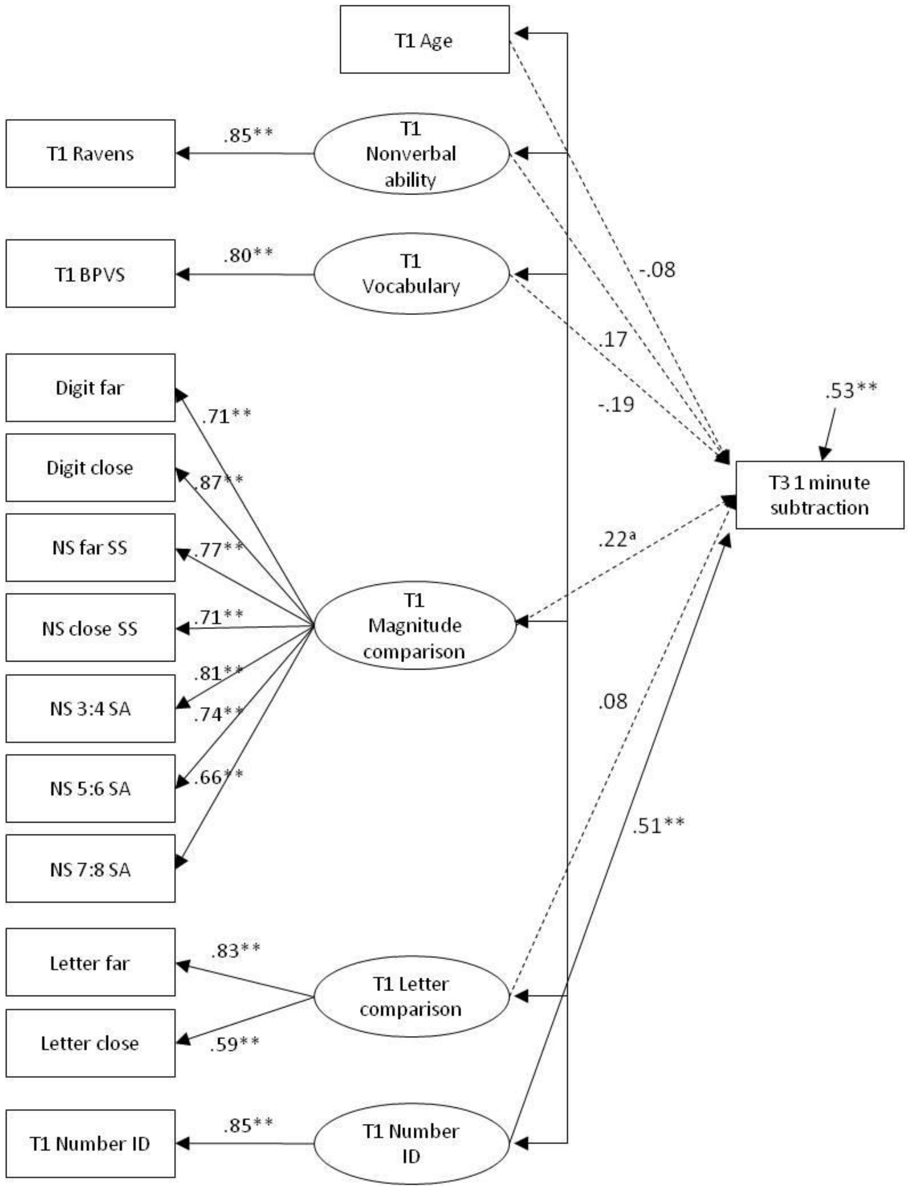


Figure 5.9. Path model predicting Time 3 one minute subtraction from Time 1 age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification. Model fit: χ^2 (61) = 107.626, *p* < .001, *RMSEA* = .066 (90% *CI* = .045-.087), *CFI* = .947, *SRMR* = .044.

[a] *p* = .069, * *p* < .05, ** *p* < .01

5.3.3.3.5. Multiplication achievement at Time 3.

Children's later multiplication ability was then predicted from the Time 1 measures. The model shown in Figure 5.10 provided an acceptable fit to the data, χ^2 (61) = 96.493, *p* = .003, *RMSEA* = .058 (90% *CI* = .035-.079), *CFI* = .959, *SRMR* = .042 (covariances presented in Appendix 22). As can be seen in Figure 5.10, there are two significant predictors of children's multiplication ability when earlier calculation ability was not controlled; children's number identification ability and magnitude comparison performance. This was the same pattern of findings as when predicting individual differences in children's later addition scores. Forty four percent of the variance in Time 3 multiplication scores was explained.

5.3.3.3.6. Summary: Predicting later calculation ability without controlling for prior calculation ability.

When children's prior calculation ability (measured using one minute addition) was not controlled in the path analyses there were some similarities but also some differences in the pattern of results. Starting with the similarities, children's earlier number identification ability was still important for later addition, subtraction and multiplication ability. Nonverbal ability, vocabulary knowledge and ability to compare letters were still not predictive of children's later addition, subtraction or multiplication scores. Number identification ability now also predicted children's subtraction achievement at Time 2 as well as Time 3. Magnitude comparison performance became a significant predictor of children's later ability to complete addition problems at both time points 2 and 3 (rather than just showing a trend towards significance at Time 3). Children's magnitude comparison performance was not a predictor of later subtraction ability, however with children's earlier calculation skill not controlled for it became a significant predictor of their ability to solve multiplication sums (alongside number identification).

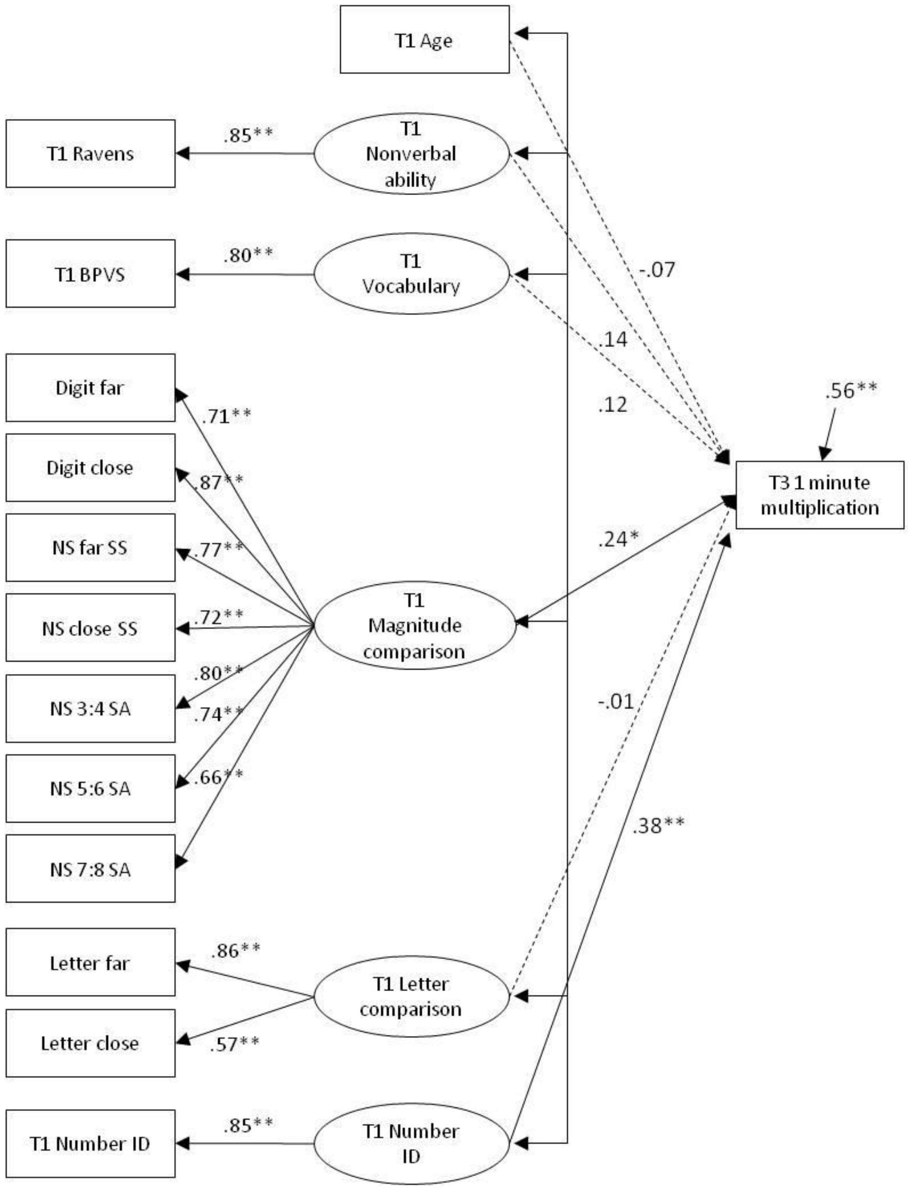


Figure 5.10. Path model predicting Time 3 one minute multiplication from Time 1 age, nonverbal ability, vocabulary, magnitude comparison, letter comparison and number identification. Model fit: χ^2 (61) = 96.493, *p* = .003, *RMSEA* = .058 (90% *CI* = .035-.079), *CFI* = .959, *SRMR* = .042.

* *p* < .05, ** *p* < .01

5.4. Discussion

5.4.1. Summary of the Study

The current chapter explored the predictors of the numerical operations addition, subtraction and multiplication, separately, using pure timed calculation tests. There were three reasons for carrying out this research: first, to extend/further the longitudinal research presented in Chapter 4 where arithmetic was assessed by a measure that included a variety of different items and children were under no time pressure to complete the test. The research presented in the current chapter used outcome measures that assessed children's ability to complete simple calculation sums under a time restriction (one minute), and therefore assessed arithmetic fluency. These types of measures have been used in previous studies (e.g. Durand et al., 2005; Holloway & Ansari, 2009) so this allows for a more comprehensive picture of what is important for children's arithmetic development. Second, it is possible that the finding in Chapter 4 that number identification was a predictor of later arithmetic achievement was due to a strong relationship and possible overlap between some of the items on the two measures. Therefore by only including calculation problems this allowed for the investigation of this suggestion. Third, it is theoretically possible that the different numerical operations depend on different precursor skills. This idea stems from the adult literature; from the proposal that there are different routes to solving different arithmetic problems, especially arithmetic facts (e.g. Dehaene & Cohen, 1997) and the body of literature that has found that individuals with brain damage can have selective impairments or preservations of different arithmetic facts (e.g. Van Harskamp & Cipolotti, 2001).

5.4.2. The Importance of Early Calculation Ability

Children's achievement on the one minute addition task was used to control for their early calculation ability in all of the analyses. The reasons behind this decision was to keep the analysis consistent across the different arithmetic operations, and because at the first time point children were not familiar with multiplication problems. As expected children's early addition skill significantly predicted children's later ability to solve addition, subtraction and multiplication problems. Children that were initially more accurate and fluent than their peers at completing addition problems continued this trend. They also performed better on the later subtraction and multiplication problems. It appears that a good grounding in the ability to solve simple addition problems is also important for solving the other arithmetic operations.

5.4.3. Number Identification as a Predictor of Later Arithmetic Skill

Children's ability to correctly match a verbally presented number to its symbolic form was found to be important for later calculation achievement; this was true for addition, subtraction and multiplication ability at Time 2 and Time 3 (with the exception of Time 2 subtraction performance). This is in line with the findings in the previous chapter when children's arithmetic was assessed using a measure that tested basic number skills (e.g. identifying and writing numbers and counting) and a mixture of calculation problems (i.e. addition, subtraction, multiplication and division). This is an important finding as it shows that the relationship between number identification ability and the WIAT Numerical Operations measure was not just reflecting a possible overlap between the items on the test that assessed basic number skills. It also shows that number identification, assessed with a measure that included not only one digit but two and three digit numbers, is important for later arithmetic achievement even when the calculation problems are relatively simple.

This finding that familiarity with symbolic numbers, and especially with numbers outside the one digit range, is important for arithmetic achievement is in line with previous literature. For example Gilmore et al. (2010) found that children's ability to name one and two digit numbers (ranging from 1 to 92) was a predictor of their arithmetic ability assessed two months later. Lembke and Foegen (2009) used a similar design of task (reading numbers ranging from 0 to 100) and found that achievement on this was related to arithmetic achievement several months later. It also provides support for the research into mathematical difficulties as it has been found that children with mathematical difficulties performed more poorly on number reading tasks than children with mathematics ability in the normal range both concurrently (e.g. De Smedt & Gilmore, 2011; Landerl et al., 2004) and longitudinally (Mazzocco & Thompson, 2005).

In Chapter 4 it was proposed that the number identification task used in the research presented in this thesis may be assessing more than children's ability to read, and familiarity with, symbolic numbers. For example, it may be tapping knowledge of place value, i.e. that the position of a digit within a multidigit number determines its value. Linked with this, the most common error that children made on the test was to choose the syntactic distractor, for example choosing 20035 instead of 235. The test could therefore also be assessing children's knowledge of the rules of the written number system, and specifically that additional zeros need to be suppressed when transcoding from the verbal to the Arabic format. It was also suggested that the relationship observed between this number identification task and arithmetic is comparable to that between letter knowledge and reading. The replication of this relationship with the arithmetic fluency measures again argues for the importance of early symbolic number knowledge and ability to map the verbal label of a number to its symbolic representation for later arithmetic achievement in general.

The only calculation measure that earlier number identification ability was not important for was subtraction ability at Time 2; in fact the only significant predictor of children's subtraction ability at Time 2 was prior addition achievement. Due to this, and the finding that number identification was a significant predictor of subtraction ability at Time 3, the data were inspected further to investigate possible reasons for these findings. The Beta weight for the importance of number identification to Time 2 subtraction was larger than for the other predictor variables (except the control of Time 1 addition) and was the closest of the measures to significance (β = .21, p = .097). There was a large range of scores on the subtraction test and the data was normally distributed. It also does not appear that children's early addition skill is explaining all of the variance in their later subtraction ability (β = .34), although when the control for prior calculation ability is removed number identification becomes the only significant predictor. It is unclear as to why the ability to correctly match a verbally presented number to its symbolic form is an important skill for all of the other operations and subtraction ability at Time 3 but not at Time 2 when prior addition skill is controlled. Children's subtraction ability, when they are 6 to 7 years of age, could therefore rely more on an earlier skill that has not been measured by this study. Children have been reported to use a variety of strategies for solving subtraction problems, for example counting, retrieval and decomposition (e.g. Farrington-Flint et al., 2009; Siegler 1989). If children are using a counting down strategy (starting from the first number and counting down the amount indicated by the number that needs to be taken away), then their ability to accurately count backwards would be important. An alternate counting strategy would be to count up to the first number from the number that needs to be taken away. While a decomposition strategy involves using knowledge of solving another sum to answer the present sum. In all of these strategies working memory may be important as children will need to hold numbers in memory whilst applying a calculation to them; in fact working memory has been found to be a predictor of later arithmetic achievement (e.g. Bull et al., 2008).

5.4.4. Magnitude Comparison Ability as a Predictor of Later Arithmetic Skill

5.4.4.1. Hypotheses.

The only other Time 1 measure in addition to Time 1 addition and number identification ability that emerged as an important skill for later calculation ability was magnitude comparison. Using Dehaene's triple code model (Dehaene & Cohen, 1995, 1997) it was hypothesised that the importance of early magnitude comparison ability would differ across the arithmetic operations. It has been proposed that there are two routes to solving arithmetic problems; a direct asemantic route where answers to simple calculation problems are retrieved directly from memory and an indirect semantic route which involves working out the answer to problems by manipulating quantities (Dehaene & Cohen, 1997). It was therefore hypothesised that magnitude comparison ability would be an important precursor skill for later addition ability, with possibly a greater influence on Time 2 than on Time 3 addition due to a change in strategy use (from needing to calculate the answer to more children being able to retrieve the answers from memory). With regards to children's subtraction achievement it was hypothesised that magnitude comparison would be an important skill at both time points (perhaps more so than for addition) due to children needing to calculate the answers and therefore manipulating quantity representations. According to the ideas put forward by Dehaene and Cohen (1997) multiplication problems would be solved using the direct asemantic route as they are over learned calculations, and therefore magnitude comparison ability would not be a predictor. Due to the age of the children at this final time point (7 to 8 years old) it was expected that not all children would be able to retrieve the solution from memory for all of the problems and that a large number of children would still need to work out the answer to at least some of the items. Therefore it is feasible that magnitude comparison would be important for later multiplication ability. However, it should be noted that this model was developed to explain adult arithmetic processing and was not suggested to explain developmental processes.

5.4.4.2. Pattern of results.

The contribution of magnitude comparison differed across the operations but not in the way expected. It was found that magnitude comparison ability was a significant predictor of children's addition scores at Time 2, with a trend towards significance for addition at Time 3; whereas it was not important for later subtraction or multiplication ability at any time point.

5.4.4.3. Addition.

Children with greater magnitude comparison ability, i.e. who were more accurate and/or faster at identifying the larger digit and the greater numerosity, gained higher addition calculation scores one year later, with the importance of this early ability decreasing with development. This finding is consistent with the idea that there are two routes to solving arithmetic problems which stems from the triple code model (Dehaene & Cohen, 1997) and the hypothesis made the that the importance of magnitude comparison ability might change over time (i.e. with a greater influence on Time 2 than on Time 3 addition) due to a change in children's strategy choice. As the children were on average seven years old at the second time point they may still need to use a calculation process to work out the answers to the problems, so this would use the indirect semantic route as it would involve manipulating quantities. As the analogue magnitude representations would be activated it is not surprising that children's ability to compare numerosities was important for later addition skill. With magnitude comparison ability only showing a trend towards significance for Time 3 addition achievement, this may reflect a change in some children's strategy choice towards being able to directly retrieve the answer from memory. According to Dehaene the direct asemantic route would be used and consequently magnitude representations would not be manipulated.

5.4.4.4. Subtraction and multiplication.

According to the ideas put forward by Dehaene and Cohen (1997) the answers to subtraction problems would be expected to be calculated using the indirect semantic route so it was hypothesised that magnitude comparison ability would be an important precursor of children's later subtraction ability. In contrast, it was found that magnitude comparison ability did not predict subtraction achievement at Time 2 or Time 3. According to Dehaene and Cohen we would not expect magnitude comparison ability to be important for later multiplication skill as it is proposed that multiplication facts are stored as verbal associations in memory. The findings from the current study do fit with this proposal as magnitude comparison ability was not found to be a predictor of multiplication achievement. However, some children at this final time point (7 to 8 years old) may not have been using a retrieval strategy on all of the items presented and may still be calculating the answer to at least some of the problems (Lemaire and Siegler, 1995). If this is true, then it would be expected that the indirect semantic route would be used and magnitude representations would be manipulated. Due to the methodology used in this current study (group testing) children's strategy use was not recorded and the actual strategies that children used are not known, this therefore warrants further investigation. In summary, while some of the data presented in this chapter are consistent with the proposals put forward by Dehaene and colleagues, some of the findings are incompatible. However, it should be noted that this model was developed to explain adult arithmetic processing and was not suggested to explain developmental processes; therefore other explanations need to be explored.

5.4.4.5. Other explanations.

Even without basing the predictions on the triple code model, children's strategy use is still an important factor to consider, and is a possible reason as to why magnitude comparison was found to be a significant predictor of children's addition skill and not their subtraction or multiplication ability. Focusing on addition, considerable heterogeneity and flexibility of strategy use has been found within and between young children (e.g. Geary & Brown, 1991; Farrington-Flint et al., 2009; Siegler, 1987). Therefore at Time 2 when children were 6 to 7 years of age some children may have been using a counting strategy to answer some of the addition problems, it is possible that they were using the more sophisticated and efficient min strategy which involves counting on from the larger number (Groen & Parkman, 1972). This would require a comparison of the two addends to identify the larger on which to start counting on from; therefore magnitude comparison would be a stage in addition. It is not surprising then that the ability to compare numerosities is important for later addition achievement. As highlighted earlier, the finding that magnitude comparison only showed a trend towards significance for Time 3 addition may be reflecting a change in some children towards a greater use of a retrieval strategy (e.g. Siegler, 1987), and therefore no need to compare the addends. With regards to subtraction, the first number in the problem is always the larger and therefore there is no need to compare the items to identify the larger. Children are first introduced to multiplication in the early school years as counting up in multiples (twos, fives and tens), later they are taught that it is the same as repeated addition and instructed to learn multiplication facts (by learning times-tables) (Department for Education and Skills, 2006). The finding that magnitude comparison ability was not a predictor of later multiplication but earlier addition ability (alongside number identification) was, suggests that early addition skill and perhaps use of this repeated addition strategy is important for multiplication achievement, rather than magnitude comparison ability.

The COMP model put forward by Butterworth and colleagues (Butterworth, Zorzi, Girelli & Jonckheere, 2001) suggests that comparison is a stage in the retrieval of addition facts. It was hypothesised that addition facts are stored in memory as max + min (i.e. 4 + 3 = 7) and that the two addends are first compared before the answer can then be retrieved. When testing the model they found that in adults, magnitude comparison reaction times predicted time to complete addition sums. They also suggest that arithmetic facts could be organised in this way due to the experience children have gained from counting on from the larger addend. From this model it would be expected that magnitude comparison ability would be important for addition achievement at both time points (whether children were using a count on from larger strategy or were able to directly retrieve the answers from memory or a mixture). It was found that magnitude comparison ability was only important for Time 2 addition achievement; it did not reach significance when predicting Time 3 addition achievement. The COMP model does therefore not fully explain the children's data in this study.

5.4.5. The Pattern of Results when Prior Arithmetic Ability is Not Controlled

5.4.5.1. Summary of the findings.

In the majority of existing studies exploring magnitude comparison as a potential predictor of children's later arithmetic achievement, prior arithmetic skill has not been assessed or controlled for. For comparison with the existing literature children's prior calculation skill was then removed from the analyses leaving magnitude comparison, letter comparison and number identification as potential predictors whilst controlling for age, vocabulary and nonverbal ability. Without controlling for prior addition achievement it was found that children's number identification ability was now a significant predictor of their later addition, subtraction and multiplication skill at both time points. Therefore children with a greater ability to correctly match a verbally presented number to its symbolic representation and understanding of place value gained higher arithmetic scores regardless of the operation, at both of the later time points. Those children with better magnitude comparison ability at Time 1 also gained higher addition calculation scores at both Time 2 and Time 3. With prior calculation skill not controlled it was also found that magnitude comparison ability predicted individual differences in children's multiplication scores but still not their subtraction achievement.

5.4.5.2. Relation to existing literature.

The finding that magnitude comparison ability was a significant predictor of later addition and multiplication ability is in line with previous literature where the relationship with arithmetic achievement in general has been explored (e.g. symbolic: De Smedt et al., 2009; Reeve et al., 2012; nonsymbolic: Libertus et al., 2013; Mazzocco et al., 2011b). However, the finding that magnitude comparison was not a predictor of subtraction ability is not. As reported previously, a lack of longitudinal relationship between nonsymbolic comparison ability (accuracy) and arithmetic has also been observed (Sasanguie et al., 2013; Sasanguie, Van den Bussche & Reynvoet, 2012). The possible reasons for a lack of relationship between magnitude comparison and arithmetic in general were discussed in Chapter 4 and could be applied here (i.e. presentation of the comparison tasks, metric used for performance on the task, statistical analyses). It is therefore worth turning to the small body of literature that has reported specific relationships with the different numerical operations. It should be noted that the focus of these papers has been on the relationship with the distance effect and not overall accuracy or speed. The findings reported here that magnitude comparison ability was a significant predictor of children's addition but not subtraction ability is consistent with some literature reporting a stronger association between the symbolic distance effect and addition than with subtraction (Landerl & Kölle, 2009; Lonnemann et al., 2011). However it is in contrast with others who found a similar strength of relationship with addition and subtraction (Vanbinst et al., 2012). With regards to the relationship between nonsymbolic comparison and addition and subtraction, both Landerl and Kölle (2009) and Lonnemann et al. (2011) reported no significant relationship with either addition or subtraction ability.

There are multiple differences with the current research that make it difficult to directly compare to the previous literature presented. First, a one factor magnitude comparison variable was used in the present analyses, as this was the best fit to the Time 1 data, rather than entering symbolic and nonsymbolic comparison separately. The three previous studies also used the distance effect rather than the overall speed with which comparisons were made and it has already been discussed earlier in this thesis that different relationships with arithmetic (and mathematics) have been found previously. The three studies reported above all have a cross-sectional design and also include older children than those involved in the current research. It has been highlighted earlier that the strategies that children use to solve simple calculation problems vary between and within individuals and this may impact upon the important contributors to achievement, although

Vanbinst et al. (2012) found similar strength correlations between the symbolic distance effect and the speed of retrieving arithmetic facts and using procedural strategies. Taken together, these points and the inconsistencies observed, warrant further investigation in order to gain a better understanding of the relationship between magnitude comparison ability and different numerical operations.

5.4.5.3. Predicting multiplication ability.

With the control for prior arithmetic ability removed from the analysis children's early magnitude comparison ability was found to predict later multiplication ability. This finding could again be linked with strategy use. In the early school years children are introduced to multiplication as counting up in multiples (twos, fives and tens), later they are taught that it is the same as repeated addition and instructed to learn multiplication facts (by learning times-tables) (Department for Education and Skills, 2006). It is feasible that at Time 3 children were using both counting and retrieval strategies (among others i.e. guessing). In the literature two repeated addition strategies have been reported to be used by children; to add the larger number the smaller amount of times (e.g. 8 x 3 = 8 + 8 + 8), and to add the smaller number the larger amount of times (e.g. 3 + 3 + 3 + 3 + 3 + 3 + 3 + 3) (see Lemaire & Siegler, 1995; Siegler, 1988). The former strategy has been found to be more accurate (Siegler, 1988), and to increase in frequency with experience of multiplication (Lemaire & Siegler, 1995). Consequently, magnitude comparison ability may therefore be important for multiplication as children are first comparing the two items of the sum in order to identify the larger (as in addition). As magnitude comparison ability was only found to be a predictor when addition skill was removed from the analysis this suggests that children's early accuracy and fluency at addition (i.e. their ability) is more important than the strategy they choose to use. This is speculation as children's strategy use was not recorded so future studies need to explore the predictors of children's multiplication ability and the strategies that they use to solve multiplication problems.

5.4.5.4. Relation to Dehaene and Cohen's proposals.

The finding that early magnitude comparison ability was not a significant predictor of children's later subtraction skill even when prior calculation achievement was excluded from the analyses speaks against the ideas put forward by the triple code model and the two routes to solving simple arithmetic problems. Especially when combined with the finding that once the arithmetic control was removed from the analysis children's magnitude comparison ability predicted individual differences in their multiplication ability.

It should be noted that this proposal by Dehaene and Cohen (1995, 1997) was to explain how single digit arithmetic problems are solved. Some of the later items on the subtraction test included subtracting a single digit number from a two digit number but only small numbers of children reached these items even at Time 3. However, this might imply that these more difficult items are even more likely to be solved by the manipulation of quantity representations and therefore via the indirect semantic route. The results presented here indicate that this adult model of solving arithmetic problems is not applicable to early development.

5.4.5.5. Possible issues with the comparison measures.

Alternatively the findings presented here could raise questions about the magnitude comparison tasks used in the current study. First, were the tasks used really a measure of an individual's analogue magnitude system? The measures used in this study were different to those typically used in the literature; they were presented to children in a paper and pencil format and they had 30 seconds to complete as many comparisons as they could. This resulted in a measure that encapsulated accuracy and speed rather than a measure of overall accuracy or the average time taken to compare two numerosities (RTs). The study carried out in Chapter 3 found that even with these differences in methodology the typical effects were observed (i.e. a distance effect was found and performance increased with development), so it appears that the measures used in this study were not so different. Second, could they just be acting as a measure of children's speed of processing or fluency at comparing items? A simple measure of speed of processing was not included in this current study so it is not known whether the relationship found with addition is reflecting this (it should be noted that not all previous studies have controlled for simple reaction time either). However, if this were true a relationship with the other operations might have been observed too. In addition a letter comparison task was included to act as a control for the comparison process and this measure was not found to be a predictor of children's calculation scores at any time point. Therefore it does not appear that the results of this study are a consequence of the different methodology applied.

5.4.5.6. Differences in results to when prior arithmetic is controlled.

When the control for earlier calculation skill is removed number identification becomes a significant predictor of Time 2 subtraction, and magnitude comparison becomes a significant predictor of Time 3 (as well as Time 2) addition and Time 3 multiplication. This raises the question of why these variables do not predict later arithmetic ability over the control. Although controlling for prior addition ability is not the same as controlling for the autoregressive effect (like in the previous chapter using the WIAT), by including a calculation control we are attempting to predict growth in arithmetic fluency rather than just later achievement. The predictor variables may therefore be important for later arithmetic achievement but do not contribute to children's growth in achievement. A second possibility is a power issue. With a predictor variable removed this would increase power. While the sample size here was large and missing data were dealt with (by Full Information Maximum Likelihood estimation), SEM is a large sample technique and ideally you would want to recruit an even greater number of participants. Nevertheless there were multiple predictors of Time 2 addition so it does not appear that the sample size was restricting the possibility of other variables being significant contributors to achievement on the other numerical operations. The advantages of using SEM techniques were reported in Chapter 2 so future studies should aim to use this more sophisticated data analysis methodology with as large a sample as possible.

5.4.6. Summary

In summary it was found that children's ability to identify the correct symbolic representation of a verbally presented number was a consistent predictor of their arithmetic ability regardless of the numerical operation. It was often as important a contributor to later arithmetic achievement as children's early addition skill. This finding supports those found in the previous chapter (Chapter 4) when arithmetic was assessed using the WIAT Numerical Operations subtest, a test that consisted of items which tapped basic numerical skills and a range of arithmetic problems that included more difficult sums (e.g. adding and subtracting two and three digit numbers). The results imply that the number identification task created for these studies is a measure of children's knowledge of symbolic (Arabic) numbers and matching the verbal representation to these. It also appears to assess children's grasp of the rules of symbolic numbers, for example, of understanding place value and the over writing rule (i.e. that when the number three hundred and fifty two is in its Arabic form the zero's to represent the hundreds and tens are dropped). Taken together the results presented in this and the previous chapter suggest that tasks that involve children's knowledge of symbolic numbers (and especially outside the one digit range) are useful measures to assess children's early arithmetic skill and to distinguish children who may fall behind their peers in their arithmetic achievement, whether this be calculation fluency or ability to answer a range of items. Future research designs should therefore include and explore further tasks that investigate children's knowledge of the symbolic number system, its rules and the mappings with verbal representations.

There was one difference to the findings in Chapter 4, magnitude comparison ability was now a significant predictor of children's addition achievement and multiplication skill (but only when the control of prior addition skill was removed). It is therefore possible that this relationship reflects the nature of the strategies used to complete the addition and multiplication problems. As mentioned previously, children use a range of strategies to answer arithmetic problems and these may differ across development. If children are indeed comparing the items to identify the larger, whether this be to choose the addend to start counting on from, to then retrieve the answer from memory, or in multiplication to add multiple times, then their ability to quickly and accurately compare the items will be important.

It should be acknowledged that the predictions for the one minute tasks are potentially rather complex. It is possible that the relationship between magnitude comparison ability and arithmetic fluency may change with development. Children who are at an early stage of their arithmetic development might be expected to show a correlation, but as children get older and become more competent they may not show this relationship since more arithmetic facts will be retrieved. The one minute tasks used in this research start with very simple problems (e.g. 2 + 1 = ?, 1 + 4 = ?, 2 – 1 = ?, 5 – 1 = ?), and on the addition task the first 20 items have a total of 10 or less (children are encouraged to develop the rapid recall of number bonds for totals up to 10 by the end of Year 2, Department for Education and Skills, 2006, this corresponds to Time 2 in the present study). Children who are less proficient at arithmetic will only complete these early items and more of these will involve fact retrieval. However, although the easier questions are listed first and children who are poorer at arithmetic will only complete these, the reason they are completing less items may be because they are using less efficient (and more time consuming) strategies such as counting rather than fact retrieval, to solve them. Children with better developed arithmetic skills may be able to retrieve the solutions to these earlier items which is why they are able to complete more items within the time limit. It is also possible that children with both poor and good arithmetic skill are using similar strategies to solve the earlier items but children who are more competent may be more efficient and faster at carrying these out. Therefore to gain a better understanding of these complex relationships more research into children's strategy use is needed, and also its inclusion as a predictor of later arithmetic is warranted.

5.4.7. Future Research Questions

These results were found by a study that included a large sample size, a careful design, and used sophisticated data analysis techniques. However, there are some remaining questions and ideas for future research. Some of the findings are different to those of existing studies, especially regarding the importance of the ability to compare magnitudes. It has been acknowledge that in order to gain a large sample to use SEM modeling techniques the methodology used by the current study differed to that typically used in the literature which could have possibly lead to these differences. Therefore in the next chapter an investigation will be carried out using a subgroup of children that have been involved in the studies presented so far in this thesis. Children's symbolic and nonsymbolic comparison ability will be investigated again using individual testing methods which will allow for more fine grained analysis. For example, measures such as the effect of distance on children's comparison times, and a Weber fraction can be calculated and then used to predict variance in children's arithmetic achievement.

Another question that remains is how reliable these magnitude comparison tasks are. There have recently been studies investigating the reliability and validity of these measures (Gilmore, Attridge & Inglis, 2011; Maloney, Risko, Preston, Ansari & Fugelsang, 2010; Price, Palmer, Battista & Ansari, 2012; Sasanguie, Defever, Van den Bussche & Reynvoet, 2011). However, these studies are all with adult participants and this therefore should be explored in children's performance.

Chapter 6.

Magnitude Comparison as a Longitudinal Predictor of Children's Arithmetic Achievement: Computerised Measures of Magnitude Comparison

6.1. Introduction

The previous chapters of this thesis have shown that number identification, children's ability to match the verbal label of a number to its symbolic form, is a consistent predictor of arithmetic achievement, both concurrently and longitudinally. This is the case when arithmetic skill is measured by an untimed test that contains items that assess children's basic understanding of number and calculation ability, and also for measures that assess calculation fluency (with the exception of subtraction ability at Time 2). In contrast, the relationship between children's magnitude comparison ability, as measured by multiple symbolic and nonsymbolic comparison tasks, and later arithmetic achievement has been inconsistent across various analyses presented thus far. The methods used in this research differed from the majority of those used in the literature: here, all measures were presented to children as paper and pencil tests in a group setting, rather than individually on a computer. Although the findings of a relationship between performance on these numerical processing tasks and arithmetic are inconsistent in the literature even when computerised measures were used, the methods used in previous chapters could still be questioned. Therefore the current study aims to investigate the longitudinal relationship between performance on magnitude comparison tasks more akin to those used in the literature and later arithmetic achievement, within the same sample of children that also completed the group based measures.

In the existing literature, the majority of the comparison tasks are presented to children (and adults) in an individual assessment session and on a computer, as a result more experimental controls can be applied and more indices of performance taken. The magnitude comparison tasks presented to children in the following studies will therefore be controlled in the following ways: the comparison tasks will be computerised and completed during individual assessment sessions. In the case of nonsymbolic comparison, stimuli with a larger range of ratio differences will be presented to enable the calculation of a Weber fraction (more detail to follow below). The stimuli will also be more tightly controlled for perceptual variables that may influence performance, i.e. size of stimuli, surface area etc. (Rousselle et al., 2004; Soltész et al., 2010). With regards to symbolic number comparison, both accuracy and time taken (reaction times) to compare the digits will be recorded and the full range of numerical distances (1 to 8) will be presented rather than just small (1 and 2) and large (5, 6, and 7) distances. This will allow for the effect of distance to be calculated as a regression slope (which reflects the decrease in time needed to compare items with each increasing distance). These computerised measures will therefore allow for multiple indices of comparison performance to be calculated separately; overall accuracy and a Weber fraction will represent performance on the nonsymbolic task and overall accuracy, mean RT, and the distance effect slope will represent performance on the symbolic version.

The Weber fraction has become an increasingly popular way to quantify an individual's performance on nonsymbolic comparison tasks but due to the design of the group based comparison tasks presented in the previous chapters, a Weber fraction could not be calculated. The Weber fraction is used as an estimate of the acuity of an individual's internal representation of number (e.g. Halberda et al., 2008; Pica et al., 2004). It is based on the findings in the literature that as the ratio between two numerosities decreases (bigger number/smaller number) discriminability also decreases; which is an example of Weber's Law. To be able to distinguish between two stimuli, they must differ by a certain amount in order for the difference to be perceived; this amount is known as the difference threshold and is operationalised as the smallest difference that can be perceived 50% of the time (Passer & Smith, 2007). This difference threshold is proportional to the magnitude of the stimuli being compared; as magnitudes increase the difference between them will also need to increase to be able to discriminate between them. This rule is known as Weber's law and this threshold can be quantified as a Weber fraction (Passer & Smith, 2007).

With regards to numerical stimuli, two psychophysical methods have commonly been used to calculate a Weber fraction. These are based on the two proposed conceptualisations of the internal number line for representing numerical magnitudes; whether it is logarithmic with fixed variability (see Dehaene, 1992) or linear with scalar variability (see Gallistel & Gelman, 1992). What is important to note is that both hypotheses give rise to equivalent behavioural patterns; discriminability decreases with decreasing ratio and with increasing magnitude. The method used in this study will be based on the logarithmic compression model using the equation proposed by Piazza et al. (2004) (for the methods based on the linear assumption see Pica et al., 2004 and Halberda et al., 2008). Both methods are said to lead to similar fits of the data (Pica et al., 2004) and propose that the Weber fraction (w) represents the precision with which numerical magnitude is represented internally; a smaller w indicating better performance and acuity.

Estimating the size of the Weber fraction (*w*) is now a prominent method of exploring numerical processing abilities across development and in relation to arithmetic achievement. It is thought that the size of the Weber fraction decreases over development (Halberda & Feigenson, 2008; Piazza et al., 2010), however, when the estimates of the Weber fraction are plotted together (see Figure 6.1) they often differ across studies and perhaps most noticeably similar estimates have been calculated for individuals with quite different ages. One possible reason for the differences across the studies could be the sample of participants included, in some studies the mean Weber fraction (*w*) reported is calculated for children with a wide age range (for example, in Libertus et al., 2011 children range in age from 2;09 to 6;01). The estimates by Halberda and Feigenson (2008), Libertus et al. (2013) and Mussolin et al. (2012) however, appear to be in close agreement over a similar age range of children, even though different methods were used by the three studies. For example, both Halberda and Feigenson, and Mussolin et al. presented numerosities as objects (e.g. cars, flags), while Libertus et al. (2013) presented stimuli as blue and yellow dots, however all three restricted the presentation time (but again by differing amounts). In contrast to this, the estimates from some of the other studies, in particular Gilmore et al. (2011), Inglis et al. (2011) and Sasanguie et al. (2013) are somewhat larger for older children (and adults). Inglis et al. suggested that a possible reason for the larger estimates attained in their study was the fact that they did not provide feedback to participants whereas Halberda and Feigenson did. Sasanguie et al. used a different method: they always presented a standard numerosity (similar to the method used by Piazza et al., 2010). Interestingly, Price et al. (2012) found that different presentation formats of the stimuli resulted in different estimates of *w*. A significantly larger *w* was found when the stimuli were intermixed (blue and yellow dots presented as one array), than when they were paired (two arrays presented next to each other, one to each side of the screen) or presented sequentially (one after the other), while the estimate from the latter two presentations did not differ. The different methods used by different research groups highlights the fact that the size of these estimates may not be directly comparable across studies.

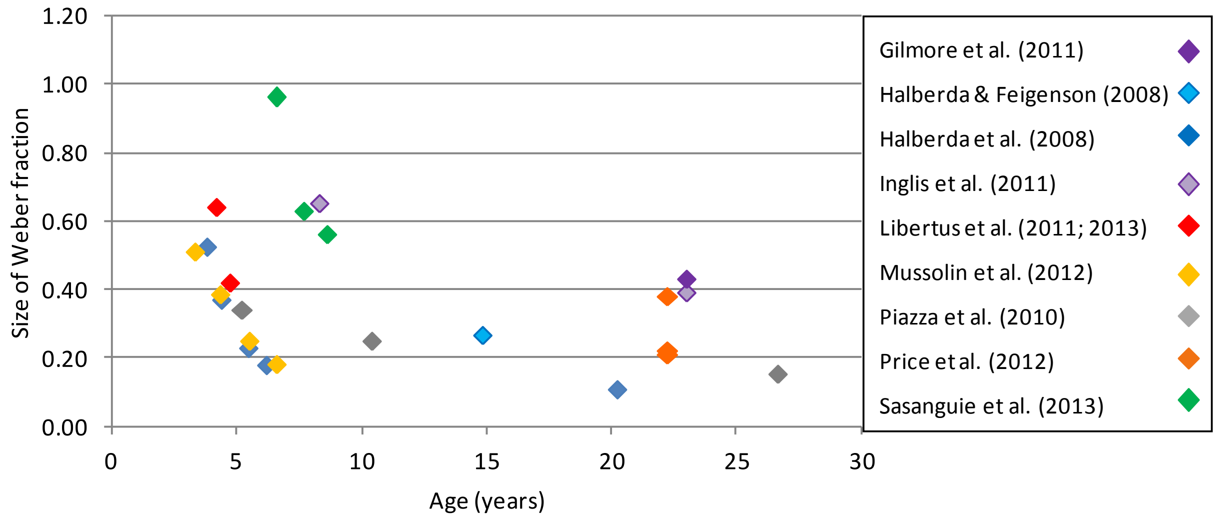


Figure 6.1. Estimates of Weber fractions for various age groups reported in the literature.

On the other hand, within the same studies the size of the Weber fraction has been found to decrease with increasing age, indicating a refinement in the acuity of numerical representations with development (Halberda & Feigenson, 2008; Libertus et al., 2013; Mussolin et al., 2012; Piazza et al., 2010; Sasanguie et al., 2013). To the best of my knowledge, only one study has investigated this longitudinally within the same sample of children, Libertus and colleagues (2013) found that on average the Weber fraction decreased in size from 0.64 at 4 years 2 months to 0.42 six months later. As these children were the same as those reported in Libertus et al. (2011) they may have covered a wide age range but this is not reported in this paper. Six months is only a relatively short period of time so more research is needed to investigate the development of the Weber fraction longitudinally. Future studies should use a narrower age range of children and the same method to estimate the size of the Weber fraction at each time point.

What these mean values reported in the literature and pictured in Figure 6.1 do not show is that large variations in the size of w have been found within the same age group. Halberda et al. (2008) found that among a group of 14 year old adolescents while the mean w was 0.27, it varied between 0.20 and 0.57. This finding has been replicated in both younger and older age groups with w varying from 0.34 to 1.17 in a group of 7 to 9 year olds, and 0.22 to 1.17 in adults (18 to 48 years, mean = 23 years) (Inglis et al., 2011). This wide range in the size of the estimates is important for investigating the relationship between individual differences in the proposed internal system for representing number (approximate number system) and arithmetic skill. As reviewed earlier (see Chapter 1) these individual differences may relate to individual differences in arithmetic achievement.

In general it has been found that those with a smaller Weber fraction have greater arithmetic and mathematics ability (Halberda et al., 2008; Inglis et al., 2011; Libertus et al., 2011, 2013; Libertus et al., 2012; Lyons & Beilock., 2011; Mazzocco et al., 2011b). Libertus et al. (2011) conducted regression analyses for overall accuracy and the Weber fraction separately, controlling for age and vocabulary knowledge. They found that speed at comparing numerosities accounted for 5% of variance and accuracy for 13% of variance in TEMA-3 scores, whereas when the Weber fraction was included instead it only accounted for 6% of variance and RT for 8%. When the same children were retested six months later accuracy on the task predicted 7% of variance in later TEMA-3 scores and RT 6% (both significant) whilst controlling for age, prior arithmetic skill (45% of variance) and vocabulary knowledge (Libertus et al., 2013). In contrast to this Sasanguie et al. (2013) found no association between the Weber fraction and either children's later arithmetic or mathematics achievement and both Inglis et al. (2011) and Price et al. (2012) found no relationship between the size of an individual's Weber fraction and concurrent arithmetic ability in adults.

There are many differences between the studies that could potentially account for the different findings, including the age range of participants, the arithmetic or mathematics outcome measure used, and the comparison task itself, with differences in the range of numerosities presented, the range of ratios tested, the presentation of the task i.e. how long the stimuli remain on the screen and whether the stimuli are intermixed or paired, and also the calculation of the Weber fraction. This all may lead to the inconsistencies observed.

An additional point which is often ignored by these studies is the reliability of nonsymbolic comparison tasks. If they lack reliability this could lead to the inconsistencies seen in the literature, in particular when they are used to predict later outcomes because the reliability will limit the size of correlations that are detectable (Amelang & Zielinski, 1997). Price et al. (2012) administered three variants of a nonsymbolic comparison task to adult participants; the stimuli were intermixed, paired or sequential. Both a Weber fraction and distance effect slope were calculated from the first half and second half of the trials (120 trials in each half). The reliability of the Weber fraction for each of the three variants was significant (intermixed: $r = .78$; paired: $r = .47$; sequential: $r = .44$), as well as the reliability of the distance effect slope (intermixed: $r = .57$; paired: $r = .78$; sequential: $r = .65$). When a delay between two administrations of the same task was applied (on average 76.39 days), Libertus et al. (2012) found that the two Weber fractions calculated were not significantly related (r = .22, p = .08) in a group of undergraduate students. In another study that was not designed to test the reliability of their measure of the ANS, Libertus et al. (2013) reported the correlations between two administrations of the task completed (by young children) on average 6.8 months apart. While there were significant associations between Time 1 and Time 2 performance, only the association between accuracy scores was moderate in strength (r = .55), with the associations between reaction times and the two Weber fractions being weak in strength (r = .22 and r = .34 respectively). Overall accuracy appears to be a more reliable measure than reaction times and the Weber fraction in children. The reliability estimates reported by these studies vary and some are weak in strength (which indicates a lack of reliability). There is also a lack of studies investigating this issue within a younger sample. Many studies are now reporting the internal consistency estimates of these measures and in general, it appears to indicate that that they are reliable (see Table 6.1) although it is possible that the large number of trials included in these nonsymbolic comparison tasks many be driving these values up. What is therefore needed to enhance the literature is a study that investigates the reliability, in particular test-retest reliability, of a nonsymbolic comparison task within a sample of children over a considerable amount of time.

Table 6.1.
Internal reliability estimates reported for nonsymbolic comparison tasks

Study	Number of trials	Reliability
Bonny & Lourenco (2013)	40	0.85 (Cronbach's alpha)
Gilmore et al. (2011)	72	Small range numerosities = 0.96
	120	Large range numerosities = 0.85 (both split half)
Halberda et al. (2012)	300	Weber fraction = 0.72 (split half)
		RTs = 0.98 (split half)
Libertus et al. (2012)	264	0.74 (split half)
Libertus et al. (2013)	60	Time 1 = 0.65, Time 2 = 0.72 (split half)
Piazza et al. (2013)	140	0.68 (split half)

In summary, as there is a lack of longitudinal research the current study explored the importance of ANS acuity as a predictor of later arithmetic achievement over a six and 18 month time lapse. The contribution of performance on a symbolic comparison task to later arithmetic ability was also investigated (a summary of the existing longitudinal literature including symbolic comparison as a predictor of arithmetic is presented in Chapter 4, Table 4.1). A narrow age range of children was included in the sample, i.e. children within the same school year, as the estimated Weber fraction could vary largely. The same arithmetic outcome measures as before were used. This all formed Study 2. To address the issue of reliability, the nonsymbolic comparison task was administered twice within the same testing session to gain an estimate of the test-retest reliability and over the multiple assessment time points to assess the stability (Study 1).

A measure of children's counting ability was also included here for two reasons; first, counting has repeatedly been reported as a significant predictor of arithmetic ability (e.g. Passolunghi et al., 2007; Reeve et al., 2012), second, to rule out the possibility that the number identification measure is capturing variance related to counting skill. In the previous analysis number identification could theoretically have been a significant predictor because of processes involved in number identification that are counting related. Both counting objects and the number identification task require children to transcode between the different representations of number as proposed in Dehaene's triple code model (Dehaene, 1992). With counting objects magnitude information must be transcoded into the verbal code, while the number identification measure requires children to map a verbal presentation of a number to its symbolic form.

6.1.1. Hypotheses

1. Consistent with the literature it is hypothesised that the size of the Weber fraction will reduce over development, indicating a refinement in children's internal numerical representations.
2. As the results observed in the existing literature as to whether the acuity of the ANS (nonsymbolic comparison ability) is related to later arithmetic achievement are inconsistent, it is both feasible that it will and it will not predict unique variance in children's arithmetic scores.
3. As children's speed at comparing symbolic digits has been found to predict individual differences in children's later arithmetic achievement it is hypothesised that this will pattern will be observed here.
4. In Chapters 4 and 5 it was found that magnitude (a combination of symbolic and nonsymbolic) comparison ability was a predictor of growth of addition ability, therefore it is hypothesised that if performance on the symbolic and nonsymbolic comparison tasks are found to predict later arithmetic achievement that it is more

likely to predict addition than the other numerical operations or achievement on the WIAT.

5. Taking the results found in the previous chapters it is predicted that children's ability to identify symbolic numbers (number identification) will remain as a significant unique predictor of children's later arithmetic achievement.
6. Counting skill at Time 1 will be an independent and unique predictor of arithmetic at Time 2 and Time 3.

6.2. Study 1

This first study had two aims, first, to investigate the reliability and stability of the nonsymbolic comparison task, calculated using overall accuracy scores and the Weber fraction, and second, to explore children's development on the task longitudinally.

6.2.1. Method

6.2.1.1. Design.

A subgroup of the sample reported in the previous chapters was selected for further testing on an individual basis. There were three time points of assessment. The first time point took place in October and November 2010 and children were at the start of school Year 2. The second time point took place between April and June 2011 which was on average 6.65 months later (range = 6 to 7 months) when children were towards the end of Year 2. The third time point took place when children were in Year 3, between March and May 2012 with an average time lapse of 10.76 months after Time 2 (range = 10 to 11 months) and 17.41 months after the first testing phase (range = 16 to 18 months).

6.2.1.2. Participants.

Individual testing consent forms were sent to parents/care givers of the children who took part in the group testing. A subgroup of children for whom consent was returned was selected to take part in individual testing sessions. As low numbers of consent forms were returned for two of the schools (11 of 46 and 13 of 22) all of the children in those schools for whom consent was received took part in the subgroup testing. In the other two schools (where 23 of 30 and 45 of 75 forms were returned) a random sample of males and a random sample of females were chosen to give a similar number of males and females overall. This strategy meant that similar numbers of children from the four schools participating were chosen.

Fifty four children were initially recruited to take part in the study. Five children in total were removed from the analyses: one child was absent when Time 2 assessments took place, one child opted out of Time 3 testing, and three children had missing data at Time 3 due to technical difficulties. The final sample included 49 children who at Time 1 had a mean age of 80.35 months (SD = 3.24), Time 2 mean age = 86.90 months (SD = 3.21) and at Time 3 were on average 97.47 months old (SD = 3.29), there were 26 males and 23 females.

6.2.1.3. Nonsymbolic comparison task.

To investigate the acuity of the internal representation of number and obtain a Weber fraction for each child a computerised nonsymbolic comparison task was administered. This task was based on the one used by Piazza et al. (2004) and was the same as that used in Sasanguie et al. (2013). The same task was used at all time points. The comparison task was presented on a Toshiba Satellite Pro L300-29D Laptop (screen size 15.4-inch, resolution 1280 x 800).

Children were presented with pairs of stimuli which appeared simultaneously. Stimuli were arrays of yellow squares within a larger grey square (450 x 450 pixels), one on the left hand side and one on the right hand side of the screen (see Figure 6.2). One of the pairs always contained the reference numerosity 16 and the second 8, 10, 12, 13, 14, 15, 17, 18, 19, 20, 22, or 24 dots to give a total of 12 conditions. There were 10 presentations of each pair with each numerosity being presented five times on both the left and the right giving a total of 120 trials. The trials were presented in a fixed random order. To avoid that performance was based on non-numerical parameters, i.e. total surface area or item size (for discussion see Dehaene, Izard & Piazza, 2005), the surface area between the two arrays presented was kept constant and the size of the squares within each stimulus varied (for more detail see Sasanguie et al., 2013). With these constraints the stimuli were created online by the program (Neurobehavioral Systems, http://www.neurobs.com).

Children were asked to choose 'which side has more squares?' and were instructed to respond as quickly and as accurately as possible but without counting the squares. They made their response by pressing either the left or right mouse button on the laptop keyboard. Stimuli remained on the screen until the child responded, with the next trial appearing after an interstimulus interval of 300ms. The task began with three practice trials and then the 120 experimental trials. If the experimenter thought that the child was taking a long time to make a decision they were reminded to make their choice without counting.

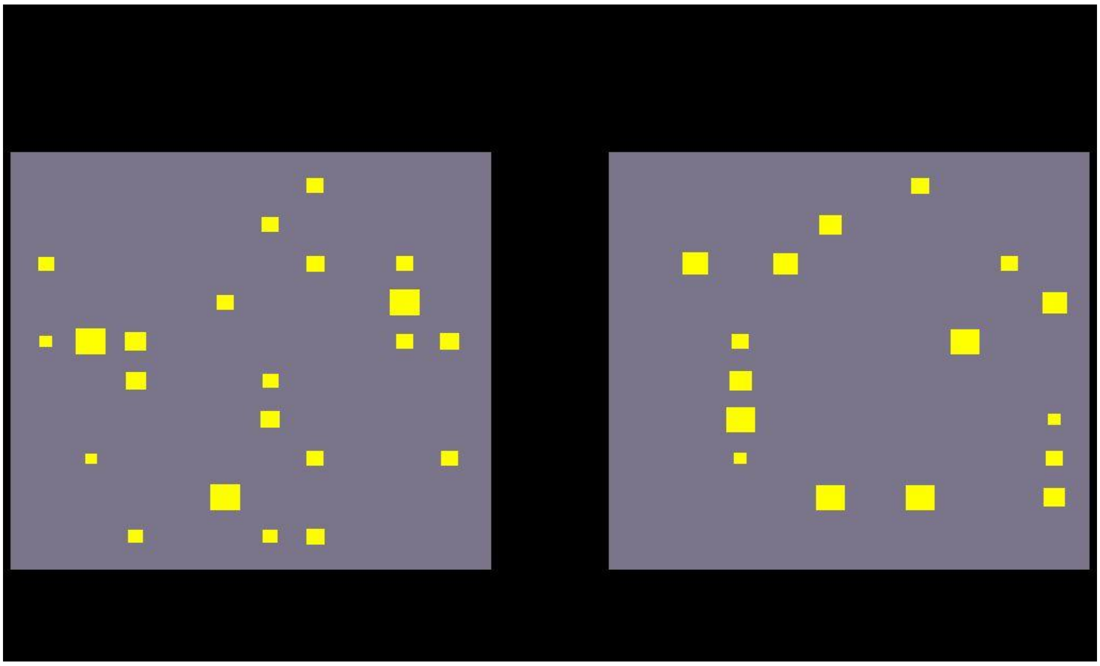


Figure 6.2. Example of the stimuli used in the nonsymbolic comparison task.

6.2.1.4. Procedure.

At all time points children were assessed individually in a quiet room or area at their school. They completed the nonsymbolic comparison task four times in total. At Time 1 they performed the task once alongside other standardised and computerised tests which are not reported here. Where possible the nonsymbolic task was administered during the second of three or four testing sessions after a symbolic comparison task (for more detail see Study 2). At Time 2 the children completed the task twice, within the same testing session (referred to as Time 2a and Time 2b). Between the two presentations of the task children completed an arithmetic test (not reported here) taking on average 5 minutes. At Time 3 children completed the comparison task again alongside other measures not reported here, the measure was the second in the testing session.

6.2.2. Results

First children's development on the nonsymbolic comparison task (or development of the ANS) will be explored, followed by an investigation into the reliability of the task. In the literature multiple ways have been used to conceptualise performance on nonsymbolic comparison tasks, including accuracy, calculating a Weber fraction, calculating the numerical distance effect and also looking at mean overall reaction times. This study investigated both children's accuracy performance and calculated a Weber fraction because those two measures are most commonly used in the longitudinal literature (e.g. Halberda et al., 2008; Libertus et al., 2013; Sasanguie et al., 2013).

6.2.2.1. Accuracy.

Mean accuracy rates showed that children's performance on the task improved over time, Time 1 = 68.76% (SD = 7.29), Time 2a = 73.62% (SD = 6.94), Time 2b = 71.62% (SD = 9.60), Time 3 = 75.03% (SD = 6.21). As children completed the task twice at Time 2 their performance on the two presentations at this time point was compared. No difference in children's accuracy between Time 2a and 2b was found ($t(48) = 1.56$, $p = .126$), therefore children's performance on the first presentation of the task at Time 2 (Time 2a) will be used in the analysis where the development on this task over time will be investigated.

The task was designed so that a given numerosity was always compared to the reference numerosity of 16 (which meant that the ratio was always X:16). Due to this design the data could also be analysed by quantifying the distance from 16 (the six distances were 1, 2, 3, 4, 6 and 8). Figure 6.3 shows that in general as the distance increased children's accuracy at comparing the items increased. To investigate this statistically a repeated measures ANOVA was performed, with both time (time points 1, 2a, and 3) and distance (1, 2, 3, 4, 6, 8) as within subject variables. Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of distance, $\chi^2(14) = 27.76$, $p = .015$ and for the interaction, $\chi^2(54) = 83.41$, $p = .007$. Therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity, ($\varepsilon = .79$ for the main effect of time and $\varepsilon = .76$ for the interaction). There was a significant effect of time on children's comparisons, $F(2, 96) = 17.83$, $p < .001$, $\eta_p^2 = .27$. Within-subjects contrasts revealed that children gained higher scores at Time 2a compared to Time 1, $F(1, 48) = 18.91$, $p < .001$, $\eta_p^2 = .28$, but that there was no significant change in accuracy between Time 2a and 3, $F(1, 48) = 1.83$, $p = .183$, $\eta_p^2 = .04$. A significant effect of distance was also observed, $F(3.96, 190.20) = 135.18$, $p < .001$, $\eta_p^2 = .74$, which confirms that children gained higher scores when the stimuli were far apart than when they were close together (all p's < .001 except between distance 3 and 4 where $p = .513$). The interaction between time and distance was not significant, $F(7.58, 363.70) = 0.88$ $p = .548$, $\eta_p^2 = .02$.

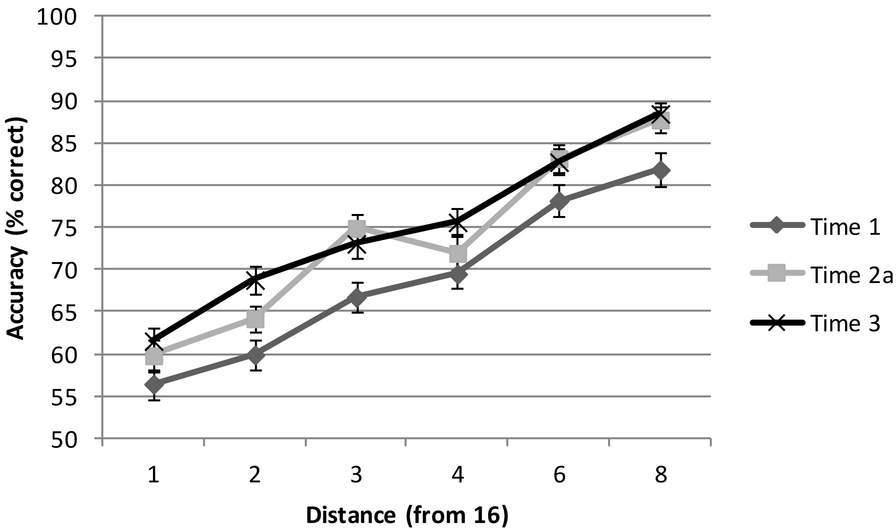


Figure 6.3. The effect of distance on children's performance on the nonsymbolic comparison tasks over time. Error bars represent standard error.

6.2.2.2. Weber fraction.

A common alternative to using accuracy scores as a metric of comparison ability is to calculate a Weber fraction (*w*). This is thought to reflect the precision with which magnitudes are internally represented (the acuity of the representations within the ANS). It was found that children's accuracy performance increased as the ratio between the two numerosities increased (bigger number/smaller number), which is in line with Weber's Law. Therefore a Weber fraction was calculated for each child individually for each time point using the psychophysical model proposed by Piazza et al. (2004). The equation in Figure 6.4 was fitted to the accuracy data with a single free parameter *w*. A smaller *w* value equates to better performance on the task and therefore reflects a more precise representation of magnitude.

$$P_{\text{larger}}(n, n_{\text{hab}}) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\log\left(\frac{n}{n_{\text{hab}}}\right)}{\sqrt{2}\,w}\right)\right)$$

Figure 6.4. Equation for fitting the Weber fraction taken from Piazza et al., 2004, p. 3 of supplementary information.

Before running any statistical analyses with the Weber fraction data the size of the Weber fraction and the fit of the model to the children's data were examined. Exclusion criteria of a *w* value larger or equal to 2.0 and a model fit (R^2 value) of less than .25 were applied. This seems appropriate given that the size of the majority of reported Weber fraction values are less than 1.0 and the model should have an acceptable fit to the data. It also meant that any outliers that may have biased the results were removed. This meant that data for seven children was removed in total, four due to *w* value of > 2.0 and a fit of R^2 <.25, and a further three due to a fit of R^2 <.25.

This sample now included 42 children who at Time 1 had a mean age of 80.26 months (SD = 3.29), Time 2 mean age = 86.79 months (SD = 3.24) and at Time 3 were on average 97.36 months old (SD = 3.35), there were 21 males and females. Mean accuracy rates were: Time 1 = 69.84% (SD = 6.04), Time 2a = 74.38% (SD = 6.48), Time 2b = 73.55% (SD = 6.74), Time 3 = 75.24% (SD = 6.16). The previous accuracy analyses were re-run with this new sample and the pattern of findings remained the same. All further analyses are run with this new sample.

As can be seen in Table 6.2 there was a large range in the size of children's Weber fractions at all time points, and the average size of the Weber fraction decreased over time. As with the accuracy data the size of children's Weber fractions from the two presentations of the task at Time 2 were compared, there was no significant difference between the two values ($t(41) = -0.63$, $p = .529$). Therefore children's performance on the first presentation of the task at Time 2 (Time 2a) will be used to investigate development of the size of the Weber fraction over time. A repeated measures ANOVA was carried out with time point as the within subjects variable. There was a significant effect of time on the size of children's Weber fractions, $F(2, 82) = 12.15$, $p < .001$, $\eta_p^2 = .23$. Within-subjects contrasts revealed that on average they decreased from Time 1 to Time 2a, $F(1, 41) = 12.73$, $p = .001$, $\eta_p^2 = .24$, but that there was no significant change between Time 2a and 3, $F(1, 41) = 1.02$, $p = .319$. This is in line with the analysis carried out on children's accuracy scores.

Table 6.2.
Descriptive statistics of the Weber fraction (w)

Time point	Measure	Mean	SD	Min	Max
Time 1	*w*	0.55	0.29	0.25	1.49
	R^2	.68	.20	.25	.95
Time 2a	*w*	0.40	0.19	0.15	0.89
	R^2	.78	.13	.41	.97
Time 2b	*w*	0.42	0.22	0.16	1.45
	R^2	.77	.17	.27	.96
Time 3	*w*	0.37	0.14	0.11	0.72
	R^2	.79	.12	.41	.98

6.2.2.3. Reliability of the nonsymbolic comparison task.

Recently there have been questions into the reliability (and validity) of nonsymbolic comparison tasks and studies investigating this have begun to emerge (see Gilmore et al., 2011; Maloney et al., 2010; Price et al., 2012; Sasanguie et al. 2011). In this current study children completed the same nonsymbolic comparison task four times over an 18 month time period with two of the assessments within the same session at Time 2. This allows for the investigation of three measures of reliability of this task: test-retest within the same time point, test-retest over an extended period of time, and internal consistency (Cronbach's α). It is important to explore the reliability of both the accuracy and the Weber fraction measures, therefore the following analyses will be run including only the children who contributed data to both the accuracy and Weber fraction analyses (n = 42).

6.2.2.3.1. Test-retest.

Simple correlations were run on children's accuracy scores (total correct) and Weber fraction values (*w*) between each of the four presentations of the task (Time 1, 2a, 2b and 3) (see Table 6.3). At Time 2, when children were on average 7 years old, they completed the nonsymbolic comparison task twice within the same session. There were significant moderate correlations between children's accuracy scores (items correct) and the two *w* values on the two presentations of the task at this time point. With regards to a longer period in time there were significant correlations between children's accuracy scores at each time point. With regards to the *w* values there were significant associations between Time 1 and Time 2a, Time 1 and Time 3 and Time 2b and Time 3; however the relationships between the *w* values at Time 1 and Time 2b, and Time 2a and Time 3 were not significant. Therefore the reliability of the Weber fraction may be weaker than the accuracy scores (total correct).

Table 6.3.
Test-retest reliability for the nonsymbolic comparison task: accuracy (above the diagonal) and Weber fraction (below the diagonal)

	Time 1	Time 2a	Time 2b	Time 3
Time 1		.38*	.34*	.45**
Time 2a	.42**		.62**	.45**
Time 2b	.26	.54**		.36*
Time 3	.31*	.29^{a}	.42**	

Note. ^{a} p .062, * $p < .05$, ** $p < .01$

6.2.2.3.2. Internal consistency.

Reliability analyses were carried out on the nonsymbolic comparison task using the data for the reduced sample (n=42). At each presentation the nonsymbolic comparison task had reasonable reliability; Time 1 Cronbach's α = .55, Time 2 a Cronbach's α = .65, Time 2b Cronbach's α = .67, and Time 3 Cronbach's α = .62, however it should be noted that it was lowest at Time 1.

6.2.3. Discussion

6.2.3.1. Children's development on the nonsymbolic comparison task.

The aim of Study 1 was to explore children's performance on a nonsymbolic comparison task over time (i.e. assess their development) and to investigate its reliability (both test-retest and internal consistency). Children were tested a total of four times, once at Time 1 when they were on average 6 years 8 months old, twice at Time 2 when they were 7 years 3 months old, and a final time (Time 3) when they were on average 8 years 1 month old. It was found that although children's accuracy at comparing nonsymbolic numerosities improved over the 18 month time period this increase was most pronounced over the initial six month period, after which it stabilised. This finding was also reflected in the Weber fractions calculated from children's accuracy scores, with the estimates reducing from 0.55 to 0.40 between the first two assessments, and then only decreasing to an average of 0.37 when children were 8 years old. This finding adds to the existing literature that the size of the Weber fraction decreases over development within the same children and provides support for the findings of Libertus et al. (2013). This suggests that this process of refinement of the approximate number systems continues through development. The non significant change between Time 2 and Time 3 suggests that this may begin to slow down between 7 and 8 years old, although this warrants further investigation before conclusions can be made. More research is therefore needed following the development of performance on nonsymbolic comparison tasks within the same children over an extended time period including children who are younger and older than those included in the current study.

6.2.3.2. Reliability.

The current study also aimed to assess the reliability of the nonsymbolic comparison measure. It was suggested earlier that this may be a possible reason behind the inconsistencies observed in the existing literature regarding the relationship (or lack of) with arithmetic achievement. Children completed the task four times in total over an 18 month time period, with two of these on the same day, so both immediate and delayed test-retest reliability (i.e. stability of performance) could be explored. The immediate test-retest correlations were significant for both the accuracy scores (as items correct) and the Weber fraction suggesting that the measure is reliable. However, while the correlations were significant between the accuracy scores at each time point, this was not true for the Weber fraction, which suggests that the estimated Weber fractions may not be stable over time. Similar results were reported by Libertus et al. (2013), who found that the estimates of Weber fractions calculated from two administrations of the comparison task completed six months apart were only weakly related, while the correlation between accuracy scores was moderate. If the longitudinal relationship with itself is not very strong then this will limit the size of correlations that are detectable with other variables (Amelang & Zielinski, 1997) which is a problem for longitudinal studies. The internal consistency of the measure was also calculated for each time point and the value obtained at Time 1 indicated relatively low reliability, especially as the number of items used to gain these estimates was high. The estimates of reliability reported here for the nonsymbolic comparison task are lower than the reliability estimates of many standardised tests, for example, the general ability measures used here. If these nonsymbolic comparison tasks that are used to gain an estimate of the precision of number representations in the approximate number system lack reliability, then this could go some way to explain the inconsistent findings reported in the existing literature of the importance of the ANS as a predictor of arithmetic.

6.3. Study 2

The overall aim of this study was to replicate the studies presented in Chapters 4 and 5. Specifically, to investigate whether the finding that number identification was a significant predictor of growth in arithmetic above the autoregressor would be replicated when comparison tasks that are more akin to those typically used in the literature are used. It could be proposed that the inconsistent findings of the importance of early magnitude comparison observed in Chapters 4 and 5 could be due to the group presented comparison tasks used. Consequently here individually presented computerised comparison tasks were used and in particular, a nonsymbolic comparison task that enabled the calculation of an individual *w*.

6.3.1. Method

6.3.1.1. Design.

The sample consisted of the children that completed the nonsymbolic comparison task at Time 1, along with other measures that will be presented below. The arithmetic data was the same as that presented in Chapters 2, 4 and 5. Both individual and group testing took place. Children were tested individually between March and July 2010 (dot counting task) this time point will be referred to as Time 1a, and then again during October and November 2010 (all other individually administered tasks; Time 1b). The group testing is as reported previously; Time 1 took place between April and July 2010, time point 2 took place between March and June 2011 and the third time point took place in February and March 2012 (see Chapter 3 for more detail). The reason for the delay between the two individual testing points at Time 1 was so that all children could be tested using the same standardised test, some of the children would have been too young to be able to complete the measures when the first assessment session took place.

6.3.1.2. Participants.

The participants were those that were involved in the longitudinal studies presented earlier in this thesis and Study 1 details how the children were recruited for the individual assessments. Fifty four children were initially recruited to take part in the study. One child was removed due to a Weber fraction (*w*) of larger than 2.0, two children due to a poor fit of the Weber fraction (i.e. R^2 <.25) and one child was removed due to incomplete symbolic comparison data (which is detailed below). The final sample included 50 children

(25 males and 25 females), their mean age at each assessment point are presented in Table 6.4.

Table 6.4.
Participant's age (in months) at each testing point.

Time point	Mean	SD	Min	Max
Individual: Time 1a	73.96	3.39	67	79
Individual: Time 1b	80.52	3.19	74	86
Group: Time 1	75.04	3.25	69	81
Group: Time 2	86.84	3.29	80	92
Group: Time 3	96.46	3.07	90	102

6.3.1.3. Assessment battery.

This study is interested in how children's performance on the measures administered at Time 1 predict their later arithmetic skill as assessed using the WIAT Numerical Operations subtest and speeded arithmetic measures. The arithmetic measures are the same as those presented in Chapters 2, 4 and 5 (a detailed description can be found on pages 51, 123, 149 respectively) so they are not detailed again here. The number identification task is also the same as reported previously and a description can be found in Chapter 2, page 52. These measures were all assessed in group presentation format. The counting, nonsymbolic comparison, symbolic comparison, verbal ability and nonverbal ability tasks were presented to children on an individual basis and are detailed below. All computerised measures were presented using a Toshiba Satellite Pro L300-29D Laptop (screen size 15.4-inch, resolution 1280 x 800).

6.3.1.3.1. Counting.

A counting task based on the one used by Landerl et al. (2004) was administered. Stimuli were arrays of randomly arranged black dots within a white square (see Figure 6.5 for an example) which appeared in the centre of the screen. The number of items ranged from one to eight. There were four trials of each numerosity with a total of 32 trials. The trials were presented in the same pseudo-random order for each child with no item occurring consecutively. Each trial was initiated by the experimenter. Children were asked to count the dots as quickly and as accurately as possible. To indicate their answer children needed to press the space bar and then recite their answer to the experimenter. The task was presented using Presentation software (Neurobehavioral Systems, http://www.neurobs.com) which recorded the reaction time for each trial, whilst the experimenter recorded the child's response. The children's answers were coded into whether or not they were correct, with the maximum score being 32.

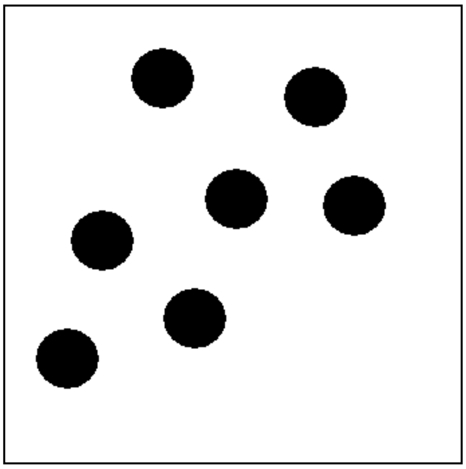


Figure 6.5. Example trial from the counting task.

6.3.1.3.2. Nonsymbolic comparison task.

The same nonsymbolic comparison task that was used in Study 1 was used again here (see section 6.2.1.3 for a detailed description).

6.3.1.3.3. Symbolic comparison task.

A classic symbolic comparison task was created. Children were presented with pairs of Arabic numerals and asked to indicate which was numerically larger. Every combination of the digits 1 to 9 (excluding ties) were presented resulting in 72 trials in total. The numerical distances ranged from 1 to 8. Each trial began with a fixation cross presented in the centre of the screen for 1000ms. The stimuli were presented in black 72 point Arial font on a white background, one on the left and one on the right hand side of the screen. Children were instructed to respond as quickly but as accurately as possible by pressing either the left or right mouse button on the laptop keypad which corresponded to the answers on the screen. Stimuli remained on the screen until the child responded. The stimuli were presented in a pseudo-random order with the same two comparison numbers never appearing successively. The task began with three practice trials and included a short break after 36 items. The software E-Prime 1.1 (Psychology Software Tools, http://www.pstnet.com) was used to present the task and record accuracy and reaction time for each trial.

6.3.1.3.4. General ability.

Individually assessed standardised ability tests were used in this study rather than the group assessed measures presented in the previous studies (Ravens Matrices and BPVS). There were two main reasons for this, first, because this study was using more tightly controlled individually presented comparison tasks, it was decided to also use general ability tests that were not reduced in item numbers and that could be standardised. Second, to check that the group measures used in the previous chapters were good estimates of children's vocabulary and nonverbal ability.

6.3.1.3.4.1. Verbal ability.

The Vocabulary subtest from the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) was used as a measure of children's verbal ability. Children were asked to provide definitions for a list of words. Testing was discontinued after five consecutive scores of 0. The maximum raw score for this age group is 56 (reliability coefficients .87 and .86 for ages 6 and 7 years respectively). The correlation with the BPVS measure used in the previous chapters was $r = .43$, $p = .003$.

6.3.1.3.4.2. Nonverbal ability.

The Matrix Reasoning subtest from the WASI was used to measure children's nonverbal reasoning skills. Children were presented with an incomplete matrix and asked to complete it by choosing the correct item from four or five options. Testing was discontinued after four consecutive scores of 0, or four scores of 0 on five consecutive items. The maximum raw score for this age group is 28 (reliability coefficients .96 and .96 for ages 6 and 7 years respectively). The correlation with the Ravens Matrices measure used in the previous chapters was $r = .78$, $p < .001$.

6.3.1.4. Procedure.

To assess arithmetic and number identification ability children were tested in whole class groups at each time point. The measures were presented alongside other tasks. Each child had their own booklet to mark their answers in and where necessary were presented with examples using the classroom whiteboard. For more detail of the presentation of the measures at Time 1 see Chapter 2 (page 57), Chapter 4 (page 124) for the WIAT Numerical Operations subtest at Time 2 and 3, and Chapter 5 (page 150) for the speeded addition measures at Time 2 and 3. With regards to the individually presented tasks (counting, nonsymbolic comparison, symbolic comparison, verbal ability and nonverbal ability) children were assessed individually in a quiet room or area within their school. All testing sessions lasted up to a maximum of 30 minutes; this depended on the child's concentration and the schools timetable. The counting task administered at Time 1a was presented as the second task in the session alongside other measures which are not reported here. At Time 1b the testing took place over three or four sessions with the Vocabulary and Matrix Reasoning subtests administered at the beginning of session one and where possible the comparison tasks were presented first in session two starting with the symbolic followed by the nonsymbolic task.

6.3.2. Results

Descriptive statistics on all measures can be seen below in Table 6.5. There was a large range of scores on all tasks. No floor or ceiling effects were observed, only one child failed to answer any items correctly and this was on the 1 minute subtraction measure at Time 2.

Table 6.5.
Descriptive statistics for all tasks (Study 2).

	N	Mean	SD	Min	Max
Verbal ability	50	22.04	5.05	9	33
Nonverbal ability	50	11.30	5.28	2	23
Nonsymbolic comparison % correct	50	70.20	6.32	55.83	82.50
Weber fraction	50	0.54	0.28	0.23	1.49
Symbolic comparison % correct	50	93.44	5.44	76.39	100.00
Symbolic comparison mean RT (ms)	50	1141.05	211.95	673.24	1607.68
Symbolic comparison slope (ms)	50	-46.51	29.41	-126.81	36.44
Counting % correct	50	94.19	6.12	75	100
Counting mean RT (ms)	50	2757.75	526.04	1865.32	4319.31
Counting mean RT 1 to 3 (ms)	50	1405.34	401.11	810.42	2970.83
Counting mean RT 4 to 8 (ms)	50	3663.68	696.22	2603.75	5268.47
Counting RT slope 1 to 3 (ms)	50	120.87	202.53	-390.63	746.50
Counting RT slope 4 to 8 (ms)	50	613.70	225.32	220.13	1076.18
Number ID	44	5.05	1.51	2	8
T1 WIAT	50	9.28	2.58	2	14
T1 1 minute addition	49	9.27	4.50	1	21
T2 WIAT	50	12.34	2.80	8	18
T2 1 minute addition	49	13.43	5.35	4	24
T2 1 minute subtraction	49	8.78	4.37	0 (1 child)	19
T3 WIAT	49	16.22	3.45	11	26
T3 1 minute addition	49	17.37	5.94	1	29
T3 1 minute subtraction	49	12.37	5.12	5	26
T3 1 minute multiplication	49	13.24	6.50	1	27

6.3.2.1. Counting.

Children's accuracy on the counting task was high (see Table 6.5) so no statistical analysis was run on this data. Reaction time data were trimmed and any trials that were +/- 3 SD's from that child's mean RT were removed (0.81 % of trials). Figure 6.6 shows that as the number of items to be counted increased so did children's RTs, $F(3.44, 168.34) = 282.87$, $p < .001$, $\eta_p^2 = .85$ (Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(27) = 158.86$, $p < .001$. Therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity, $\varepsilon = .49$).

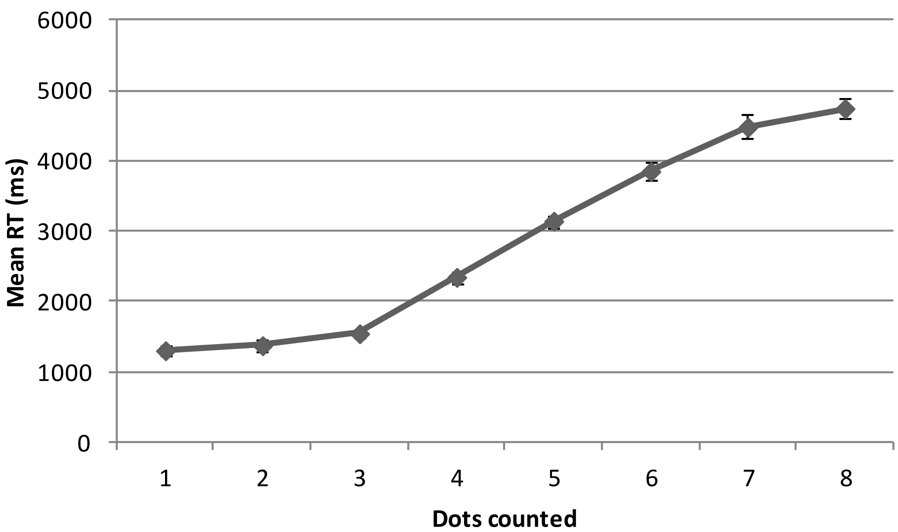


Figure 6.6. Dot counting reaction times (in ms). Error bars represent standard error.

When counting one to three items there was only a small increase in the children's RTs per extra item, whereas when counting four or more items there appears to be a greater increase in the time needed. This is reflected in the mean RTs for these two sets of numbers (see Table 6.5). To investigate this further, as in Landerl et al. (2004), the best fitting regression lines were calculated for each child based on their RTs for the items within what could be termed the subitizing range (1 to 3) and the counting range (4 to 7). Reaction time on the task is therefore predicted by the number of items and the gradient of the slope thus reflects the time increase to count an additional item. The descriptive statistics for the regression slopes are reported in Table 6.5 and it can be seen that there is a steeper slope (i.e. greater increase in the time needed to count an extra item) for items in the counting range than in the subitizing range. The two slopes differed significantly from each other, $t(49) = -10.56$, $p < .001$.

6.3.2.2. Nonsymbolic comparison.

The pattern of children's performance on the nonsymbolic comparison task with this sample of children remained the same as found in Study 1 (overall accuracy rate = 70.20% (SD = 6.32) compared to 68.76%). As the distance between the numerosities increased so did children's accuracy, $F(5, 245) = 44.76$, $p < .001$, $\eta_p^2 = .48$.

6.3.2.3. Symbolic comparison.

For the symbolic comparison task children's accuracy and reaction time (in ms) was recorded for each trial. To analyse the reaction time (RT) data only correct trials were included and a data trimming procedure was applied removing any trial that was +/- 3 SD's from that child's mean RT (1.25% of trials removed). In general, accuracy on symbolic comparison tasks is high therefore the analysis will concentrate on children's reaction time data (accuracy results can be found in Table 6.5). Figure 6.7 shows the typical pattern observed in the literature that children's reaction times decrease as the distance between the digits increases.

To classify the effect of distance a linear regression was calculated on each individual's reaction times, predicting reaction time on the task from numerical distance (see De Smedt et al., 2009; Fias et al., 1996; Lorch & Myers, 1990). The gradient of the slope reflects the size of the distance effect, with steeper slopes indicating larger distance effects. As previous research has found that children's RTs increase as the distance decreases, it would be expected that the relationship, and therefore the slope, would be negative. There are several advantages to using this method to quantify the effect of distance over others as pointed out by Fias et al. (1996): it allows the quantification of the size of the effect rather than testing for the presence or absence of the effect (as in an ANOVA), the increase in distance is considered as a continuous variable rather than just selecting distances to reflect 'small' and 'large' distances (as in the distance effect calculation), and by calculating a regression slope for each individual this reduces the chance of overestimating the distance effect due to group averaging.

The mean regression slope was found to be negative and there was a large range in the gradient of the children's slopes (see Table 6.5). Three children were found to have positive slopes but were still included in the analyses. A one-sample t-test revealed that the slope was significantly different from zero, $t(49) = -11.184$, $p < .001$, which indicates a significant effect of numerical distance on children's comparison times.

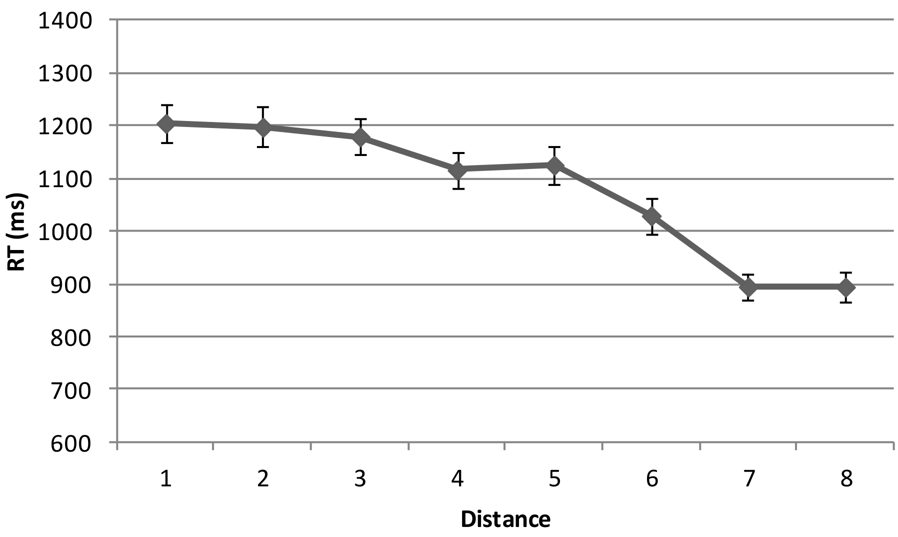


Figure 6.7. The effect of distance on children's reaction times (in ms) on the symbolic comparison task. Error bars represent standard error.

6.3.2.4. General ability.

There was a large range in children's ability scores on both tasks (see Table 6.5). Children's Full scale IQ standard scores ranged from 80 to 134. It was decided to include all children in further analyses so that there was a wide range of abilities (it should be noted that no child was registered with special educational needs at their school).

6.3.2.5. Correlation analyses.

The aim of this study was to investigate the relationship between the proposed measures of magnitude representation and later arithmetic achievement and to replicate the studies presented in Chapters 4 and 5 but with tasks more akin to those used in the literature. First, the associations between the possible predictors of children's later arithmetic achievement were examined and are presented in Table 6.6 (simple and partial correlations controlling for age at the first individual testing phase). In general the strength of the relationships were similar whether controlling for age or not. Second, the relationships (partial correlations controlling for age) between these possible predictor variables and the arithmetic measures were examined (Table 6.7).

With regards to children's performance on the comparison tasks, there was no association between children's accuracy and speed on the symbolic comparison task indicating that there was no speed-accuracy trade-off. As expected there was a very strong negative correlation between children's mean accuracy on the nonsymbolic comparison task and the Weber fraction, indicating that children who were more accurate on the task had smaller Weber fractions. This is not surprising, because the calculation of the Weber fraction is based entirely on the accuracy data. There were weak to moderate significant associations between children's mean accuracy on the two comparison tasks. No significant correlations were found between any of the comparison measures and number identification. There was no significant association between children's accuracy and speed on the dot counting task indicating that there was no speed-accuracy trade-off. There were also no significant correlations between any of the comparison and the counting measures but there were significant moderate relationships between number identification and children's mean RT and slope calculated using 4 to 8 items.

Concentrating on the relationships between the possible predictors of arithmetic and arithmetic, both of the general ability measures were significantly associated with the arithmetic tasks, the only exception is Vocabulary and Time 2 one minute subtraction. The correlations were stronger between arithmetic and the non verbal measure, than with vocabulary knowledge. There were no significant relationships between any of the measures of magnitude comparison and arithmetic. It should be noted however, that there were marginal but weak associations between one minute multiplication at Time 3 and the size of children's Weber fraction and their accuracy on the symbolic task.

With regards to the counting tasks there were no significant associations between children's accuracy performance and arithmetic. There were also no relationships with children's arithmetic and counting 1 to 3 items (mean RT and slope). Children's overall mean RT was significantly associated with Time 1 and 2 speeded addition, Time 2 WIAT, and speeded subtraction at Time 3. Children's mean RT when only including items in the counting range (4 to 8) was also significantly associated with these arithmetic measures and speeded addition and multiplication at Time 3 (the associations with overall mean RT were marginal). There were weak but significant correlations between achievement on the WIAT at time points 1 and 2 and the slope calculated using 4 to 8 items. All of the significant relationships were negative indicating that children who were faster at counting gained higher arithmetic scores on these measures.

In contrast to this, children's number ID ability was significantly and positively related to every arithmetic measure that the children completed, this is true for every time point, the strength of these associations range from moderate to strong. Therefore children who were better at matching a spoken number to its symbolic form at Time 1 had higher arithmetic scores at all time points. This is not surprising given that these were the same results presented earlier but only including the reduced subgroup.

Table 6.6.

Simple (above the diagonal) and partial (below the diagonal) correlations controlling for age between the possible predictor variables.

	Vocabulary	Matrix reasoning	NS accuracy	NS Weber fraction	Symbolic accuracy	Symbolic mean RT	Symbolic slope (RT)	Dot counting accuracy	Dot counting mean RT	Dot counting mean RT 1 to 3	Dot counting mean RT 4 to 8	Dot counting slope (RT 1 to 3)	Dot counting slope (RT 4 to 8)	Number ID
Age (T1 individual)	.23	.15	-.17	.04	.05	-.02	-.10	-.06	-.09	.04	-.13	-.17	-.06	.23
Vocabulary		.43**	.18	-.27	.19	-.19	-.22	-.02	-.12	.11	-.18	-.07	-.17	.44**
Matrix reasoning	.42**		.01	-.02	.17	-.07	.21	.09	-.12	.09	-.19	-.22	-.25	.54**
NS accuracy	.23	.03		-.89**	.35*	.19	-.10	-.12	-.06	-.04	-.03	-.01	-.09	.02
NS Weber fraction	-.29*	-.03	-.90**		-.45**	-.15	.21	-.17	-.01	-.03	-.03	.19	.02	-.07
Symbolic accuracy	.19	.16	.37*	-.46**		.09	-.35*	-.18	.11	.03	.01	-.20	.01	.20
Symbolic mean RT	-.19	-.07	.19	-.15	.09		.08	-.02	.06	-.14	.14	-.01	.17	-.24
Symbolic slope (RT)	-.20	.23	-.12	.22	-.35*	.08		.13	-.08	-.01	-.12	.13	-.18	.05
Dot counting accuracy	-.01	.10	-.14	.17	-.17	-.02	.13		.23	.08	.13	.02	-.05	.11
Dot counting mean RT	-.10	-.10	-.07	-.00	.01	.06	-.09	.22		.72**	.94**	.15	.50**	-.26
Dot counting mean RT 1 to 3	.10	.08	-.04	-.03	.03	-.14	-.01	.08	.73**		.52**	-.01	.05	.09
Dot counting mean RT 4 to 8	-.16	-.17	-.05	-.03	.02	.14	-.13	.12	.96**	.53**		-.19	.63**	-.38*
Dot counting slope (RT 1 to 3)	-.04	-.20	-.04	.20	-.19	-.02	.12	.01	-.17	-.01	-.21		-.19	.02
Dot counting slope (RT 4 to 8)	-.17	-.25	-.11	.02	.02	.17	-.19	-.05	.50**	.06	.63**	-.20		-.40**
Number ID	.41**	.52	.06	-.08	.19	-.24	.08	.13	-.24	.08	-.36*	.06	-.39**	

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

Table 6.7.
Partial correlations controlling for age between the possible predictor variables and arithmetic.

	T1 WIAT	T1 1 minute addition	T2 WIAT	T2 1 minute addition	T2 1 minute subtraction	T3 WIAT	T3 1 minute addition	T3 1 minute subtraction	T3 1 minute multiplication
Vocabulary	.33*	.36*	.47**	.44**	.19	.50**	.43**	.37**	.37**
Matrix reasoning	.52**	.45**	.53**	.61**	.49**	.58**	.44**	.58**	.50**
NS accuracy	.02	.07	.04	.03	-.04	.16	.20	.13	.23
NS Weber fraction	-.01	-.16	-.07	-.05	.05	-.16	-.17	-.14	-.27[a]
Symbolic accuracy	.11	.22	.22	.16	.11	.26	.16	.25	.28[b]
Symbolic mean RT	-.15	-.08	-.20	-.05	-.19	-.12	-.09	-.02	-.19
Symbolic slope (RT)	.22	.10	.04	.21	.17	-.05	.14	.06	.03
Dot counting accuracy	.12	.06	.03	.19	-.01	-.01	.04	.06	.09
Dot counting mean RT	-.16	-.37*	-.42**	-.36*	-.07	-.18	-.28[c]	-.33*	-.27[d]
Dot counting mean RT 1 to 3	.02	-.15	-.09	-.14	.09	-.03	-.11	-.23	-.06
Dot counting mean RT 4 to 8	-.22	-.41**	-.51**	-.44**	-.13	-.20	-.30*	-.33*	-.33*
Dot counting slope (1 to3)	-.05	-.05	.11	-.10	-.01	-.03	.05	-.03	-.06
Dot counting slope (4 to 8)	-.29*	-.23	-.36*	-.26	-.17	-.10	-.09	-.17	-.17
Number ID	.55**	.48**	.62**	.68**	.60**	.58**	.73**	.68**	.66**

Note. [a] $p = .069$, [b] $p = .057$, [c] $p = .058$, [d] $p = .062$
* $p < .05$, ** $p < .001$

6.3.2.6. Regression analyses.

To investigate whether any of the variables predicted variance in children's later arithmetic achievement, a series of hierarchical multiple regressions were computed for each arithmetic measure separately. Due to the relatively small sample size a maximum of four variables were entered into each model at a time. To reduce the number of variables being entered into the analyses (and therefore increase power) a composite score to represent children's general ability was created by averaging z-scores for the Vocabulary and Matrix Reasoning subtests. To control for age, all raw scores were first residualised for age at the time they were assessed and these standardised residuals were used in the analyses.

6.3.2.6.1. Predicting arithmetic achievement from general ability, magnitude comparison and number identification achievement.

There were two aims of the first set of regression analyses: first, to investigate whether children's performance on the magnitude comparison tasks predicted variance in children's later arithmetic scores (although the lack of significant correlations between the measures is acknowledged) and second, and more importantly, to explore whether the findings in the previous chapters that number identification is important for arithmetic development could be replicated once age, general ability and magnitude comparison ability were controlled. Therefore general ability was entered as Step 1, followed by children's accuracy on the nonsymbolic comparison task and the slope (distance effect) on the symbolic comparison task (Step 2). Children's accuracy on the nonsymbolic task was chosen instead of the Weber fraction to represent performance on the task as there was a very strong correlation between the two metrics and the reliability of the accuracy measure was slightly better (see Study 1, section 6.2.2.3.1, page 199). No significant relationships were found between any of the symbolic comparison measures and arithmetic. The slope was chosen because it gives information about the effect of distance on children's comparisons rather than just how fast they were (accuracy on the task was not used as it was high). Number identification (number ID) was entered as the final step (Step 3) because there were moderate to strong correlations with the arithmetic measures and to investigate whether it would prove to be a unique predictor of variance in later arithmetic achievement in addition to the variables entered in earlier steps.

6.3.2.6.1.1. Predicting Time 2 arithmetic achievement.

The models predicting Time 2 WIAT Numerical Operations (Table 6.8), one minute addition (Table 6.9) and one minute subtraction (Table 6.10) achievement were all significant. A similar pattern of results emerged whether arithmetic at Time 2 was measured by a task that assessed basic number skills and calculation ability or calculation fluency. Neither of the nonsymbolic or symbolic magnitude comparison tasks predicted significant variance in children's later arithmetic ability. However, it should be noted that when predicting children's speeded subtraction scores, in the final step the contribution of the symbolic slope (distance effect) almost reached significance (p = .051). In contrast, children's early ability to identify numbers was found to be a significant unique predictor of their later arithmetic achievement in all models, with the variance explained ranging from 10% (for the WIAT) to 22% (for subtraction). Nonverbal ability was a unique predictor of variance in children's later Numerical Operations and addition scores but not their subtraction scores.

It should be noted that results were the same when the Weber fraction was used instead of accuracy on the task and when the symbolic slope was replaced with mean accuracy and mean RT, apart from the Time 2 subtraction results. When the symbolic slope was replaced with mean accuracy or mean RT the contribution from this task was no longer marginally significant.

Table 6.8.
Hierarchical regression analysis predicting Time 2 WIAT Numerical Operations achievement from general ability, nonsymbolic comparison, symbolic comparison and number ID.

Step	Variable	Beta	*t*	R^2 change	Total R^2
1	General ability	.65**	5.52	.42**	.42
2	General ability	.65**	5.28	.00	.42
	NS accuracy	-.06	-0.46		
	Sym slope (RT)	-.01	-0.08		
3	General ability	.43**	3.12	.10**	.52
	NS accuracy	-.07	-0.65		
	Sym slope (RT)	-.01	-0.06		
	Number ID	.38**	2.86		

Note. $F(4, 43)$ 10.71, $p < .001$, $R^2 = .52$. * $p < .05$, ** $p < .01$.

Table 6.9.
Hierarchical regression analysis predicting Time 2 one minute addition achievement from general ability, nonsymbolic comparison, symbolic comparison and number ID.

Step	Variable	Beta	*t*	R^2 change	Total R^2
1	General ability	.68**	5.87	.46**	.46
2	General ability	.67**	5.55	.01	.47
	NS accuracy	-.09	-0.78		
	Sym slope (RT)	.05	0.38		
3	General ability	.41**	3.22	.13**	.60
	NS accuracy	-.11	-1.08		
	Sym slope (RT)	.07	0.61		
	Number ID	.45**	3.57		

Note. $F(4, 42) = 14.31$, $p < .001$, $R^2 = .60$. * $p < .05$, ** $p < .01$

Table 6.10.
Hierarchical regression analysis predicting Time 2 one minute subtraction achievement from general ability, nonsymbolic comparison, symbolic comparison and number ID.

Step	Variable	Beta	*t*	R^2 change	Total R^2
1	General ability	.40**	2.82	.16**	.16
2	General ability	.35*	2.43	.06	.22
	NS accuracy	-.09	-0.62		
	Sym slope (RT)	.23	1.57		
3	General ability	.03	0.17	.22**	.44
	NS accuracy	-.11	-0.92		
	Sym slope (RT)	.25[a]	2.01		
	Number ID	.57**	3.83		

Note. $F(4, 42) = 7.45$, $p < .001$, $R^2 = .44$. [a] $p = .051$. * $p < .05$, ** $p < .01$.

6.3.2.6.1.2. Predicting Time 3 arithmetic achievement.

To investigate whether the same pattern of findings would hold over a longer time period children’s arithmetic achievement at Time 3 (when children were 7 to 8 years old) was predicted from the same measures assessed at Time 1. As before the models predicting WIAT Numerical Operations (Table 6.11), one minute addition (Table 6.12) and one minute subtraction (Table 6.13) achievement were all significant, as was the model predicting children’s one minute multiplication (Table 6.14). The pattern of findings did indeed hold, as before neither of the magnitude comparison measures were significant predictors of children's later arithmetic achievement. However, the symbolic distance effect (slope) was no longer a marginally significant predictor of subtraction ability at this later time point. Number identification was again found to be a unique predictor of children's later WIAT Numerical Operations (7% of variance), addition (26% of variance), subtraction (19% of variance) and multiplication (22% of variance) ability. General ability was only a unique predictor of WIAT Numerical Operations and not any of the timed arithmetic measures.

Table 6.11.
Hierarchical regression analysis predicting Time 3 WIAT Numerical Operations achievement from general ability, nonsymbolic comparison, symbolic comparison and number ID.

Step	Variable	Beta	*t*	R^2 change	Total R^2
1	General ability	.59**	4.72	.35**	.35
2	General ability	.63**	4.86	.03	.38
	NS accuracy	-.01	-.07		
	Sym slope (RT)	-.17	-1.28		
3	General ability	.45**	2.99	.07*	.45
	NS accuracy	-.02	-.19		
	Sym slope (RT)	-.15	-1.22		
	Number ID	.32*	2.19		

Note. $F(4, 42) = 7.70$, $p < .001$, $R^2 = .45$. * $p < .05$, ** $p < .01$.

Table 6.12.
Hierarchical regression analysis predicting Time 3 one minute addition achievement from general ability, nonsymbolic comparison, symbolic comparison and number ID.

Step	Variable	Beta	*t*	R^2 change	Total R^2
1	General ability	.52**	3.92	.27**	.27
2	General ability	.52**	3.69	.00	.27
	NS accuracy	.03	0.22		
	Sym slope (RT)	.02	0.15		
3	General ability	.16	0.18	.26**	.54
	NS accuracy	.00	0.02		
	Sym slope (RT)	.05	0.44		
	Number ID	.62**	4.63		

Note. $F(4, 42) = 10.95$, $p < .001$, $R^2 = .54$. * $p < .05$, ** $p < .01$.

Table 6.13.

Hierarchical regression analysis predicting Time 3 one minute subtraction achievement from general ability, nonsymbolic comparison, symbolic comparison and number ID.

Step	Variable	Beta	t	R^2 change	Total R^2
1	General ability	.56**	4.28	.31**	.31
2	General ability	.56**	4.10	.00	.31
	NS accuracy	.00	0.02		
	Sym slope (RT)	-.02	-0.14		
3	General ability	.25	1.79	.19**	.50
	NS accuracy	-.02	-0.18		
	Sym slope (RT)	.01	0.05		
	Number ID	.53**	3.85		

Note. $F(4, 42) = 9.62$, $p < .001$, $R^2 = .50$. * $p < .05$, ** $p < .01$.

Table 6.14.

Hierarchical regression analysis predicting Time 3 one minute multiplication achievement from general ability, nonsymbolic comparison, symbolic comparison and number ID.

Step	Variable	Beta	t	R^2 change	Total R^2
1	General ability	.48**	3.54	.23**	.23
2	General ability	.48**	3.33	.01	.24
	NS accuracy	.09	0.65		
	Sym slope (RT)	.03	0.21		
3	General ability	.15	1.03	.22**	.40
	NS accuracy	.07	0.54		
	Sym slope (RT)	.06	0.46		
	Number ID	.57**	3.92		

Note. $F(4, 42) = 8.10$, $p < .001$, $R^2 = .46$. * $p < .05$, ** $p < .01$.

It should be noted that the results were the same when the Weber fraction was used instead of accuracy on the task and when the symbolic slope was replaced with mean accuracy or mean RT.

6.3.2.6.2. Predicting arithmetic from general ability, counting and number identification.

It is clear from the analysis so far that children's early ability to identify numbers is an important contributor to their later arithmetic achievement. It is possible that this task is capturing variance related to counting skill. The next phase of the analysis will investigate whether counting ability is a predictor of children's later arithmetic achievement, and more importantly, whether individual differences in number identification still predict variance in children's arithmetic scores once counting ability is controlled. Due to the findings that neither of the magnitude comparison tasks were significant predictors of children's arithmetic, this step was removed. Separate models were run for each arithmetic measure with general ability entered in Step 1, children's mean RT for 4 to 8 items in Step 2, and finally number identification in Step 3. The mean RT for the counting range was chosen to represent counting ability because it showed the highest correlation (of all the counting metrics) with arithmetic achievement.

6.3.2.6.2.1. Predicting Time 2 arithmetic achievement.

The models predicting Time 2 WIAT Numerical Operations (Table 6.15), one minute addition (Table 6.16) and one minute subtraction (Table 6.17) achievement were all significant. As can be seen in Table 6.15 and Table 6.16 general ability, counting speed and number identification ability were all significant unique predictors of children's arithmetic achievement at Time 2, as measured by the WIAT Numerical operations subtest and the timed addition test, accounting for a total of 60% and 66% of variance respectively. Early number identification ability accounted for 5% of unique variance in children's later WIAT achievement and 7% of variance in addition skill once general ability and counting ability were controlled.

The model predicting children's subtraction ability accounted for 37% of the variance (Table 6.17), with the only significant unique predictor being number identification which accounted for 21% of the variance in children's scores. Children's counting speed did not make a unique contribution even before number identification was entered, once general ability was controlled (Step 2).

Table 6.15.
Hierarchical regression analysis predicting Time 2 WIAT Numerical Operations achievement from general ability, counting and number ID.

Step	Variable	Beta	*t*	R^2 change	Total R^2
1	General ability	.65**	5.52	.42**	.42
2	General ability	.56**	5.23	.13**	.55
	Counting	-.38**	-3.50		
3	General ability	.42**	3.45	.05*	.60
	Counting	-.31**	-2.93		
	Number ID	.28*	2.23		

Note. $F(3, 43)$ 20.29, $p < .001$, $R^2 = .60$. * $p < .05$, ** $p < .01$.

Table 6.16.
Hierarchical regression analysis predicting Time 2 one minute addition achievement from general ability, counting and number ID.

Step	Variable	Beta	*t*	R^2 change	Total R^2
1	General ability	.68**	5.87	.46**	.46
2	General ability	.59**	5.68	.14**	.60
	Counting	-.38**	-3.72		
3	General ability	.42**	3.72	.07**	.66
	Counting	-.30**	-3.03		
	Number ID	.33**	2.78		

Note. $F(3, 42) = 25.57$, $p < .001$. $R^2 = .66$. * $p < .05$, ** $p < .01$

Table 6.17.
Hierarchical regression analysis predicting Time 2 one minute subtraction achievement from general ability, counting and number ID.

Step	Variable	Beta	*t*	R^2 change	Total R^2
1	General ability	.40**	2.82	.16**	.16
2	General ability	.39*	2.65	.00	.17
	Counting	-.04	-0.28		
3	General ability	.10	0.65	.21**	.37
	Counting	.10	0.73		
	Number ID	.58**	3.57		

Note. $F(3, 42) = 7.64$, $p < .001$, $R^2 = .37$. * $p < .05$, ** $p < .01$.

6.3.2.6.2.2. Predicting Time 3 arithmetic achievement.

As before the regression analyses were repeated for predicting arithmetic achievement at Time 3 when the children were 7 to 8 years old. All models were significant and the amount of variance accounted for in arithmetic scores ranged from 43% (for the WIAT) to 56% (for speeded addition). A similar pattern emerged for all the Time 3 arithmetic outcome measures. Number identification was found to be a significant unique predictor of children's later WIAT scores (Table 6.18; 6% of variance), speeded addition (Table 6.19; 20% of variance), speeded subtraction (Table 6.20; 15% of variance) and speeded multiplication (Table 6.21; 17% of variance) once general ability and counting speed were controlled. Children's general ability was also found to make a unique contribution to achievement on the WIAT.

The ability to identify numbers at Time 1 was still found to predict variance in all of the arithmetic measures at each time point when mean counting speed for 4 to 8 items was replaced with each of the different metrics to quantify counting ability (overall accuracy, overall mean RT, mean RT for 1 to 3 items, slope for 1 to 3 items, slope for 4 to 8 items).

Table 6.18.
Hierarchical regression analysis predicting Time 3 WIAT Numerical Operations achievement from general ability, counting and number ID.

Step	Variable	Beta	*t*	R^2 change	Total R^2
1	General ability	.59**	4.72	.35**	.35
2	General ability	.57**	4.36	.01	.37
	Counting	-.12	-0.94		
3	General ability	.41**	2.76	.06*	.43
	Counting	-.05	-0.38		
	Number ID	.31*	2.06		

Note. $F(3, 42) = 9.73$, $p < .001$, $R^2 = .43$. * $p < .05$, ** $p < .01$.

Table 6.19.
Hierarchical regression analysis predicting Time 3 one minute addition achievement from general ability, counting and number ID.

Step	Variable	Beta	*t*	R^2 change	Total R^2
1	General ability	.52**	3.92	.27**	.27
2	General ability	.45**	3.46	.08*	.36
	Counting	-.30*	-2.28		
3	General ability	.17	1.28	.20**	.56
	Counting	-.17	-1.48		
	Number ID	.56**	4.22		

Note. $F(3, 42) = 16.41$, $p < .001$, $R^2 = .56$. * $p < .05$, ** $p < .01$.

Table 6.20.
Hierarchical regression analysis predicting Time 3 one minute subtraction achievement from general ability, counting and number ID.

Step	Variable	Beta	*t*	R^2 change	Total R^2
1	General ability	.56**	2.82	.31**	.31
2	General ability	.50**	2.65	.05[a]	.36
	Counting	-.24[a]	-0.28		
3	General ability	.25[a]	0.65	.15**	.52
	Counting	.13	0.73		
	Number ID	.49**	3.57		

Note. $F(3, 42) = 13.90$, $p < .001$, $R^2 = .52$. [a] $p = .072$. * $p < .05$, ** $p < .01$.

Table 6.21.
Hierarchical regression analysis predicting Time 3 one minute multiplication achievement from general ability, counting and number ID.

Step	Variable	Beta	*t*	R^2 change	Total R^2
1	General ability	.48**	3.54	.23**	.23
2	General ability	.42**	3.10	.07[a]	.30
	Counting	-.26[a]	-1.93		
3	General ability	.16	1.11	.17**	.47
	Counting	-.14	-1.51		
	Number ID	.52**	3.57		

Note. $F(3, 42) = 11.60$, $p < .001$, $R^2 = .47$. [a] $p = .061$. * $p < .05$, ** $p < .01$.

6.3.2.6.3. Predicting arithmetic while controlling for prior arithmetic skill.

The SEM analyses in Chapters 4 and 5 were run both with and without controlling for children's prior arithmetic skill. For completeness and to investigate whether children's early ability to match a verbally presented number to its symbolic form (number ID) is a significant predictor once children's prior arithmetic is controlled, the analyses were run with an arithmetic control. Separate models were run for each arithmetic measure with general ability entered in Step 1, children's Time 1 WIAT Numerical Operations or one minute addition were entered in Step 2, and finally in Step 3 number identification. Children's score on the one minute addition test was used as a control for prior arithmetic skill for all of the speeded measures (as in Chapter 5).

6.3.2.6.3.1. Predicting Time 2 arithmetic achievement.

As can be seen in Table 6.22, the model accounted for 59% of variance in WIAT Numerical Operations scores. Alongside children's prior arithmetic skill, the only other significant unique predictor of children's arithmetic achievement was their general ability at Time 1. The contribution of children's early number identification ability once age, general ability and the autoregressor were controlled was marginally significant.

Table 6.22.
Hierarchical regression analysis predicting Time 2 WIAT Numerical Operations achievement from general ability, prior arithmetic achievement and number ID.

Step	Variable	Beta	*t*	R^2 change	Total R^2
1	General ability	.65**	5.04	.42**	.42
2	General ability	.41**	2.77	.13**	.55
	WIAT	.43**	3.54		
3	General ability	.31*	2.40	.04[a]	.59
	WIAT	.35*	2.69		
	Number ID	.26[a]	1.97		

Note. $F(3, 43) = 19.32$, $p < .001$, $R^2 = .59$. [a] $p = .056$. * $p < .05$, ** $p < .01$.

Children's early number identification ability was a significant unique predictor of their later addition (Table 6.23) and subtraction (Table 6.24) achievement, even when controlling for age, general ability and prior addition skill, explaining 5% and 13% of variance respectively. General ability was also a unique predictor of children's addition, but not subtraction ability.

Table 6.23.

Hierarchical regression analysis predicting Time 2 one minute addition achievement from general ability, prior addition achievement and number ID.

Step	Variable	Beta	t	R^2 change	Total R^2
1	General ability	.68**	5.87	.46**	.46
2	General ability	.39**	3.76	.24**	.69
	1 minute addition	.57**	5.53		
3	General ability	.26*	2.50	.05**	.75
	1 minute addition	.49**	4.98		
	Number ID	.29**	2.56		

Note. $F(3, 42) = 38.04$, $p < .001$, $R^2 = .75$. * $p < .05$, **$p < .01$.

Table 6.24.

Hierarchical regression analysis predicting Time 2 one minute subtraction achievement from general ability, prior addition achievement and number ID.

Step	Variable	Beta	t	R^2 change	Total R^2
1	General ability	.40**	3.81	.16**	.16
2	General ability	.18	2.30	.14**	.31
	1 minute addition	.44**	2.58		
3	General ability	-.02	1.27	.13**	.43
	1 minute addition	.32*	1.84		
	Number ID	.45**	2.50		

Note. $F(3, 42) = 9.88$, $p < .001$, $R^2 = .43$. * $p < .05$, **$p < .01$.

6.3.2.6.3.2. Predicting Time 3 arithmetic achievement.

Number identification was not found to be a significant predictor of children’s achievement on the Numerical Operations test at Time 3 once prior arithmetic skill was controlled (Table 6.25). In fact only prior arithmetic achievement was a significant unique predictor (although the contribution of general ability was marginal). When predicting children’s calculation fluency at Time 3, number identification remained a unique predictor of children’s speeded addition (Table 6.26), subtraction (Table 6.27), and multiplication (Table 6.28). With general ability and prior calculation skill controlled, number identification accounted for 22% of the variance in addition scores, 13% in subtraction scores, and 19% in multiplication scores.

Table 6.25.
Hierarchical regression analysis predicting Time 3 WIAT Numerical Operations achievement from general ability, prior WIAT achievement and number ID.

Step	Variable	Beta	t	R^2 change	Total R^2
1	General ability	.59**	4.25	.35**	.35
2	General ability	.35*	1.96	.12**	.47
	WIAT	.42	3.08		
3	General ability	.28[a]	1.44	.03	.49
	WIAT	.34*	2.29		
	Number ID	.21	1.64		

Note. $F(3, 42) = 12.65$, $p < .001$, $R^2 = .49$. [a] $p = .066$. * $p < .05$, ** $p < .01$.

Table 6.26.
Hierarchical regression analysis predicting Time 3 one minute addition achievement from general ability, prior addition achievement and number ID.

Step	Variable	Beta	t	R^2 change	Total R^2
1	General ability	.52**	3.20	.27**	.27
2	General ability	.38*	1.81	.05	.33
	1 minute addition	.27	2.12		
3	General ability	.14	0.37	.22**	.54
	1 minute addition	.11	1.02		
	Number ID	.59**	4.55		

Note. $F(3, 42) = 15.31$, $p < .001$, $R^2 = .54$. * $p < .05$, ** $p < .01$.

Table 6.27.
Hierarchical regression analysis predicting Time 3 one minute subtraction achievement from general ability, prior addition achievement and number ID.

Step	Variable	Beta	t	R^2 change	Total R^2
1	General ability	.56**	4.28	.31**	.31
2	General ability	.35*	2.44	.11**	.42
	1 minute addition	.40**	2.82		
3	General ability	.16	1.13	.13**	.55
	1 minute addition	.27*	2.07		
	Number ID	.45**	3.34		

Note. $F(3, 42) = 16.00$, $p < .001$, $R^2 = .55$. * $p < .05$, ** $p < .01$.

Table 6.28.
Hierarchical regression analysis predicting Time 3 one minute multiplication achievement from general ability, prior addition achievement and number ID.

Step	Variable	Beta	t	R^2 change	Total R^2
1	General ability	.48**	3.54	.23**	.23
2	General ability	.36*	2.29	.04	.27
	1 minute addition	.23	1.42		
3	General ability	.14	0.90	.19**	.46
	1 minute addition	.08	0.52		
	Number ID	.55**	3.66		

Note. $F(3, 42) = 10.95$, $p < .001$, $R^2 = .46$. * $p < .05$, ** $p < .01$.

Where children's counting ability was a significant unique predictor of children's arithmetic ability (Time 2 WIAT and Time 2 1 minute addition; Section 6.3.2.6.2), the previous analysis was repeated with counting ability also controlled. This will investigate whether children's ability to identify numbers is a unique predictor of arithmetic once age, general ability, prior arithmetic skill, and counting ability are controlled. Separate models were run for the two arithmetic measures with general ability entered in Step 1, children's Time 1 WIAT Numerical Operations or one minute addition were entered in Step 2, counting (Mean RT for 4 to 8 dots) in Step 3, and finally in Step 4 number ID.

It was found that children's number identification ability was not a significant unique predictor of their Time 2 WIAT Numerical Operations achievement when age, general ability, prior arithmetic skill, and counting ability were controlled (Table 6.29). However, number identification remained as a significant unique predictor (4% of variance) of children's speeded addition achievement (Table 6.30).

Table 6.29.

Hierarchical regression analysis predicting Time 2 WIAT Numerical Operations achievement from general ability, prior WIAT achievement, counting and number ID.

Step	Variable	Beta	*t*	R^2 change	Total R^2
1	General ability	.65**	5.52	.42**	.42
2	General ability	.41**	3.29	.13**	.55
	WIAT	.43**	3.47		
3	General ability	.37**	3.26	.10**	.65
	WIAT	.37**	3.30		
	Counting	-.33**	-3.33		
4	General ability	.31*	2.56	.02	.67
	WIAT	.32*	2.71		
	Counting	-.29**	-2.96		
	Number ID	.17	1.42		

Note. $F(3, 42) = 10.95$, $p < .001$, $R^2 = .67$. * $p < .05$, **$p < .01$.

Table 6.30.

Hierarchical regression analysis predicting Time 2 one minute addition achievement from general ability, prior WIAT achievement, counting and number ID.

Step	Variable	Beta	*t*	R^2 change	Total R^2
1	General ability	.68**	5.87	.46**	.46
2	General ability	.39**	3.76	.23**	.69
	1 minute addition	.57**	5.53		
3	General ability	.39**	3.98	.04*	.73
	1 minute addition	.47**	4.37		
	Counting	-.22*	-2.31		
4	General ability	.28**	2.75	.04*	.77
	1 minute addition	.42**	4.13		
	Counting	-.17[a]	-1.90		
	Number ID	.25*	2.50		

Note. $F(4, 42) = 31.32$, $p < .001$, $R^2 = .77$. [a] $p = .066$. * $p < .05$, **$p < .01$.

6.3.3. Discussion

The aim of Study 2 was to investigate whether comparison tasks more akin to those used in the existing literature (i.e. individually presented and computerised) would consistently predict children's later arithmetic achievement and whether number identification would predict arithmetic once performance on these comparison tasks was controlled. Like in the previous chapters, number identification was found to be a significant unique predictor of both types of arithmetic measure (timed and untimed), even when controlling for age, magnitude comparison performance and counting speed (introduced as a possible confound). It even remained as a unique predictor of children's calculation fluency (timed arithmetic) when children's prior addition skill was controlled. In contrast, neither the ability to compare nonsymbolic numerosities nor symbolic numbers was related to children's later arithmetic achievement. These findings will be discussed below.

6.3.3.1. Nonsymbolic comparison as a predictor.

Even though the expected ratio effect was observed in the nonsymbolic comparison task, neither children's overall accuracy nor the Weber fraction was significantly related to their achievement on any of the arithmetic measures, six or 18 months later. In order to be comparable to the large group analyses and to observe whether any relationships were present once individual differences in children's general ability were controlled, performance on the nonsymbolic comparison task was included in the regression analyses predicting children's later arithmetic achievement. Accuracy on the nonsymbolic comparison task was chosen to represent the possible contribution of the ANS (over the Weber fraction) because it was found to be a more reliable measure than the Weber fraction. Consistent with the correlation analysis, accuracy on the comparison task was not found to be a significant predictor of variance in arithmetic scores, whether this was six or 18 months later, and whether arithmetic was assessed by a timed or untimed measure.

These results support a recent study by Sasanguie et al. (2013) who reported no significant relationship between either accuracy on the comparison task or the Weber fraction and achievement on either a timed arithmetic test (calculation fluency) or untimed mathematics measure. They are however, in contrast to those reported by Libertus et al. (2013) and Mazzocco et al. (2011b). Both studies found that performance on the nonsymbolic comparison task was related to later arithmetic achievement, in addition

Halberda et al. (2008) and Mazzocco et al (2011a) found a significant retrospective relationship.

One clear difference between the current study (and also Sasanguie et al.'s (2013) study) and the longitudinal studies where a relationship was found, is the age range of the participants. Libertus et al. (2013) and Mazzocco et al. (2011b) both assessed nonsymbolic comparison ability in a younger age group of children (on average two years younger) than those assessed in the current study. In two other longitudinal studies (Halberda et al., 2008; Mazzocco et al., 2011a) the relationship was actually investigated *retrospectively*, with the Weber fraction being calculated for adolescents and relating this to prior arithmetic achievement. There are also studies with adult participants where a concurrent association between performance on a nonsymbolic comparison task and arithmetic has not been found (Castronova & Göbel, 2012; Inglis et al., 2011; Price et al., 2012). Overall, this suggests that the association between ANS acuity and arithmetic (or mathematics) may not be the same across development. It is possible that the association between the acuity of the ANS and arithmetic may only be present earlier in development and possibly before formal schooling commences, while later the retrospective association could be reflecting a refinement in the numerical representations of the ANS due to better arithmetic competence earlier in development (Piazza et al., 2013). The present study explored the longitudinal relationship between the ANS and arithmetic over two years with children aged between 6 and 8 years. It is not able to give answers about the relationship in younger or older children, or adults, therefore further research is needed investigating this relationship from a much earlier age (before children being formal schooling) and covering an extended period of time as children progress through numeracy education.

The lack of association found in the current study along with the inconsistent findings reported in the literature pose questions about the reliability and validity of these nonsymbolic comparison measures. As reported earlier, the reliability of these measures is far from ideal which questions its use in longitudinal designs. Recently studies have been published investigating the convergent validity of these measures; in particular whether different versions/presentations of the nonsymbolic task are measuring the same underlying cognitive construct. To the best of my knowledge only one study so far has calculated a Weber fraction from different variants of nonsymbolic comparison task (Price et al., 2012). Price et al. (2012) reported significant convergent validity of the Weber fraction and numerical distance effect (although not all of the distance effects were significantly related). This replicated the findings of Maloney et al. (2010) who used smaller numerosities and two different paradigms to calculate the numerical distance effect. On the other hand, Gilmore et al. (2011) found no relationship between small (1 to 9) and large (9 to 70) numerosity comparison tasks. Together, the inconsistent results found across different versions of nonsymbolic comparison task, low reliability and mixed findings when investigating the validity of these measures raise the question about whether these tasks are actually measuring the acuity of the ANS at all. More research is therefore needed in order to clarify this, particularly into the reliability and validity of the different formats of nonsymbolic comparison task and into whether there are differences in the association with arithmetic between the different tasks within the same sample. There is also a lack of research exploring these issues in children. This is important given that these magnitude comparison tasks and especially nonsymbolic variants are proposed to assess the acuity of basic representations of number, which are in turn suggested to be the foundations for learnt arithmetical (and mathematical) knowledge to build upon. The present research has aimed to begin to address this but future work is needed.

6.3.3.2. Symbolic comparison as a predictor.

The acuity of the approximate number system, assessed with a nonsymbolic comparison task, did not significantly predict children's later arithmetic achievement. There is however more evidence from the existing literature of a relationship between symbolic number comparison and arithmetic. Therefore at Time 1 children also completed a symbolic comparison task that was based on those typically seen in the literature (e.g. De Smedt et al., 2009). Overall their accuracy on the task was high so the analysis focused on the speed of their comparisons. In order to quantify the effect of distance on children's comparisons a regression slope was calculated, the average slope was found to be negative which is consistent with findings reported that as the distance between numbers increases, reaction times decrease (e.g. De Smedt et al., 2009; Moyer & Landauer, 1967). In contrast to some previous literature, no significant associations were found between this measure of the distance effect and children's later arithmetic achievement (e.g. De Smedt et al., 2009; Reeve et al., 2012). However, this finding is in line with two recent studies by Sasanguie and colleagues who found no relationship between the distance effect (calculated as a slope or difference score) and children's later arithmetic achievement (Sasanguie, Van den Bussche & Reynvoet, 2012; Sasanguie et al., 2013).

Although no association with arithmetic achievement was found, for completeness it was included in the regression analyses exploring the predictors of children's arithmetic achievement. Consistent with the correlation analysis, and reflecting the findings for the nonsymbolic comparison task, individual differences in the symbolic distance effect were not found to predict significant variance in later arithmetic scores, providing support for the findings of Sasanguie, Van den Bussche and Reynvoet (2012) and Sasanguie et al. (2013).

In contrast to previous literature (De Smedt et al., 2009; Sasanguie, Van den Bussche and Reynvoet, 2012; Sasanguie et al., 2013), no significant longitudinal relationships were found with accuracy on the task or mean reaction times. The design of the current study was very similar to that of De Smedt et al. (2009); the children were of a similar age range (6 years 4 months compared to 6 years 8 months in the current study) and the comparison task and method of analysing the data were also the same. In the studies by Sasanguie, Van den Bussche and Reynvoet (2012) and Sasanguie et al. (2013) children of the same age were included but their age range was wider (5 to 7 years and 6 to 8 years old respectively) and the measures used to represent speed at comparing the digits also differed (RT adjusted to also reflect error and median RT respectively). However, it is not thought that this would make a substantial difference to the results.

Desoete et al. (2012) used a different type of symbolic comparison measure and found that its predictive relationship with arithmetic differed depending on the time lag; no relationship was observed with arithmetic achievement assessed one year later but an association was found two years later (although some of the significant correlations were still weak in strength). In the current study a trend towards significance between children's accuracy on the symbolic comparison task and their multiplication scores at Time 3 was observed, with children with poorer symbolic comparison accuracy gaining lower scores on the multiplication task. As there are relatively few longitudinal studies exploring the association between symbolic number processing and later arithmetic achievement and the findings again appear to be inconsistent, then further research is needed to clarify this.

6.3.3.3. Number identification as a predictor.

In contrast to the findings of a lack of relationship between the magnitude comparison tasks and arithmetic, number identification was consistently found to be related to later arithmetic achievement. This was the case for each arithmetic measure at each time point, with children who were better able to match the verbal label of a number to its symbolic form gaining higher arithmetic scores almost one and two years later. This is perhaps not surprising as these children were a subsample of those reported in the previous chapters and this is therefore the same data, however, even with age, children's general ability and performance on the individually presented magnitude comparison tasks controlled, number identification was a significant predictor of later arithmetic scores. The amount of unique variance explained ranged from 7% of individual differences in Time 3 WIAT scores to 26% of variance in Time 3 addition scores. To investigate whether the number identification task was capturing variance related to counting skill, children's speed (mean RT) at counting 4 to 8 items was included in the regression analyses in place of the magnitude comparison tasks. The same findings remained, with number identification being a significant unique predictor of later arithmetic achievement; this was the case for the standardised untimed arithmetic test and each of the different calculation fluency measures (with variance explained ranging from 5% of Time 2 WIAT scores to 22% of Time 2 subtraction scores).

This finding that familiarity with symbolic numbers is important for arithmetic achievement is consistent with existing literature (e.g. De Smedt & Gilmore, 2011; Gilmore et al., 2010; Landerl et al., 2004; Lembke & Foegen, 2009; Mazzocco & Thompson, 2005). As discussed previously, the number identification task may not have only assessed children's knowledge of the symbolic (Arabic) number system but may have also tapped their transcoding ability, mapping between different representations of number, and understanding of place value. The ability to map one number representation onto another appears to be an important skill for later arithmetic success (e.g. Kolkman, Kroesbergen & Leseman, 2013: Krajewski & Schneider, 2009) and recently it has been found that numerical knowledge (number reading and knowing the quantity that the symbol represents) is a mediator between number skills (e.g. counting ability, knowledge of counting rules, concepts of more and less) assessed at preschool and formal mathematics ability assessed one year later (Purpura, Baroody & Lonigan, 2013). The findings observed in this study support this.

The most common error children made on the number identification task was to choose the syntactic distractor over the correct form (i.e. they failed to suppress the extra zero's), which is also a common error children make when transcribing the verbal number to its digital code (e.g. Power & Dal Martello, 1990; Seron & Fayol, 1994). This is interesting as Moeller et al. (2011) also found that in German speaking children the number of errors they made when transcoding one, two and three digit numbers in first grade was related to the number of errors they made on an addition task two years later. These measures appear to assess children's understanding of the place value system and this is therefore important for arithmetic development. The results of the current study extend the findings to English speaking children.

In order to be comparable to the large group analyses and to assess the importance of number identification as a predictor of *growth* in arithmetic skill, the autoregressor was then included in the analyses. Number identification remained as a significant unique predictor of calculation fluency even when prior addition skill (along with age and general ability) was controlled. Furthermore, number identification was a better predictor of children's calculation ability at Time 3 (addition, subtraction and multiplication) than either their general ability or prior addition skill. In contrast, when predicting arithmetic growth using the WIAT Numerical Operations subtest, the contribution of early number identification ability only showed a trend towards significance for achievement one year later and was not a unique predictor of Time 3 arithmetic ability. This suggests that children's early knowledge of the symbolic number system may be even more important for calculation fluency skills where quick access to the quantity representations of numbers is required.

6.3.3.4. Counting as a predictor.

A measure of children's counting ability was also included in this study for two reasons; first, counting has repeatedly been reported as a significant predictor of arithmetic ability (e.g. Passolunghi et al, 2007; Reeve et al., 2012), and second, to rule out the possibility that the number identification measure was capturing variance related to counting skill. In general, the finding that counting is a significant longitudinal predictor of arithmetic achievement was replicated (e.g. Koponen, Salmi, Eklund & Aro, 2013; Passolunghi et al, 2007; Reeve et al., 2012; Aunola, Leskinen, Lerkkanen & Nurmi, 2004).

The counting measure chosen assessed children's counting speed and the typical pattern seen in the literature was observed, with the time needed to count small numbers of items (1 to 3) increasing at a much smaller rate than for larger numbers of items (4 to 8) (e.g. Landerl et al., 2004; Reeve et al., 2012). Children may therefore have been subitizing rather than actually counting these small numbers of items (Kaufman et al., 1949). Neither the average time nor the slope calculated from items within the subitizing range was related to children's arithmetic ability, whereas the average time and the slope calculated when items were in the counting range were related to children's Time 2 WIAT and addition (fluency) achievement and Time 3 calculation fluency (but not Time 2 subtraction skill and Time 3 WIAT).

Since the average time to count items within the counting range was significantly correlated with arithmetic performance, it was included in the regression analyses. Counting speed predicted significant variance in Time 2 WIAT achievement and both Time 2 and Time 3 addition ability, there was also a trend towards significance for subtraction and multiplication skill at the later time point. Children who were quicker at counting gained higher arithmetic scores later. However, when number identification ability was included in the analyses, counting remained as a unique predictor only for Time 2 WIAT and addition achievement. There are two main points to take from this: number identification is not capturing variance related to counting skill as it contributed additional variance when predicting every arithmetic measure once counting speed was controlled. Second, counting speed appears to be important for Time 2 arithmetic achievement (however, not subtraction skill) but is less important for Time 3 arithmetic, perhaps when other skills are more important (i.e. working memory, strategy use).

Number identification ability remained as a significant predictor of later arithmetic achievement even when counting speed was controlled. This was the case even for the timed arithmetic measures assessing calculation fluency, which could be described as simpler arithmetic tasks than the WIAT Numerical Operations test. This provides further support for the proposal that a good grounding in the knowledge of the symbolic number system (and links to the verbal number representation) is important for children's arithmetic development.

A possible reason that counting speed appears to be more important for Time 2 than Time 3 arithmetic achievement could be a shift in strategy use. Children may have been more likely to use a counting strategy more often at Time 2 than Time 3, and at Time 3 more children may have been able to rely on a fact retrieval strategy to answer some of the addition and multiplication problems (Carpenter & Moser, 1984; Farrington-Flint et al., 2009; Groen & Parkman, 1972; Lemaire & Siegler, 1995). Therefore children who were faster at counting at Time 1 were also faster with applying a count on strategy in the arithmetic assessments. With regards to the change in predictive relationship with the WIAT arithmetic measure, at Time 3 children will have needed to use more than their counting ability to answer some of the more difficult problems presented, they will have also needed to draw on the strategies they have been taught in order to be able to solve them.

6.3.3.5. How do the results compare to the large group analysis?

In order to be able to compare the findings from the large group design (Chapters 4 and 5) and small group design (current chapter) studies, similar steps were taken. Both age and children's general cognitive ability were controlled, symbolic and nonsymbolic comparison tasks were completed by the children, and the same data reflecting their ability to identify symbolic numbers and arithmetic achievement was used. In both the small and large group analysis, children's early number identification ability was found to be a significant predictor of the individual differences in later arithmetic achievement scores. This was true for both untimed and timed (calculation fluency) measures of arithmetic. However, it should be noted that when children's prior arithmetic skill was controlled the contribution of number identification ability did not reach significance for Time 2 subtraction in the large group study and Time 3 WIAT in the small group study.

There were however, some differences regarding the importance of numerical processing ability. In the large sample SEM analyses using group administered magnitude (symbolic and nonsymbolic) comparison tasks, children's performance on these measures was found to be important for later calculation fluency skill (again with the exception of Time 2 subtraction achievement). When children's earlier achievement on the addition task was included to control for prior arithmetic skill, magnitude comparison ability remained only as a significant predictor of their addition ability at Time 2, and showed a trend towards significance for the same task another year later. In contrast, neither performance on the computerised symbolic nor nonsymbolic comparison task was found to be a unique predictor of children's later arithmetic achievement. However, in the large group analyses where magnitude comparison was found to be a significant predictor its contribution was always less than number identification. In summary, using two different methods to assess children's numerical processing abilities and two different methods of analyses led to very similar findings: overall, children's knowledge of the symbolic number system measured using a number identification task, was a better predictor of their later arithmetic achievement than the measures designed to assess the acuity of numerical representations.

6.3.3.6. Summary.

In summary, children's achievement on the number identification measure at Time 1 was important for their later arithmetic achievement assessed almost one and two years later. It is possible that this task measures children's knowledge of the symbolic number system, the ability to transcode between the different number representations and understanding of place value, all of which appear to be a crucial foundation for more complex arithmetic knowledge to build upon. In contrast, children's numerical processing ability, assessed using different presentations of magnitude comparison task, appears to be inconsistently related to children's later arithmetic ability. In addition to this, the findings from Experiment 1 show less than ideal estimates of reliability for the task typically used to assess the acuity of the ANS (nonsymbolic comparison). Together with the inconsistent findings in the literature, this suggests that measures of numerical processing/underlying numerical representations may not be the best measures to use when investigating predictors of children's arithmetic development. Future research should therefore be focused on exploring the relationship between number identification tasks and later arithmetic achievement; further investigation is needed to gain a clearer idea of what part of this number identification task is important for arithmetic development.

Chapter 7.
General Discussion

The overall aim of this thesis was to explore the predictors of children's arithmetic development with a specific focus on magnitude comparison ability. An individual's performance on symbolic and nonsymbolic comparison tasks is thought to be an indicator of the acuity of their internal representation of number (e.g. Halberda & Feigenson, 2008; Sekuler & Mierkiewicz, 1977). Furthermore it has been suggested that this is the foundation upon which formal arithmetic knowledge builds (Ansari & Karmiloff-Smith, 2002; Barth et al., 2005). Within the numerical processing literature there still remains a lack of longitudinal studies that aim to explore this proposal. This thesis therefore set out to investigate this hypothesis in a group of children who were at an early stage of formal schooling (Year 1) and followed them for almost two years (21 months; Year 3). A novel feature of this research was to assess children in whole class groups in order to recruit a sample large enough to use sophisticated data analysis techniques (structural equation modeling; Chapters 2, 4 and 5), while also assessing a subsample of children individually with computerised measures and on standardised tests (Chapter 6). This thesis also aimed to address a limitation in the existing literature by exploring children's development on the magnitude comparison tasks within the same group of children (Chapters 3 and Chapter 6 Study 1).

Chapter 2 aimed to assess the underlying latent factors that different comparison tasks (Arabic digit, nonsymbolic, letter) may have in common and to investigate the concurrent relationships with children's arithmetic achievement. The study also aimed to induce the effects of distance and ratio on comparison performance (performance increases when the distance and ratio between items increases) in novel group measure tasks. It was found that this novel methodology did indeed produce the intended effects, with an effect of distance being observed on children's digit, nonsymbolic and letter comparisons, and an effect of ratio found on nonsymbolic comparisons. In a CFA, the best fitting solution to the data showed that measures that involved comparing magnitudes (symbolic digits and nonsymbolic numerosities) loaded on the same factor, with letter comparison as a separate factor. All three tasks had a very similar presentation and task demands so it appears that the symbolic digit and nonsymbolic comparison tasks are tapping the same underlying construct, possibly an individual's internal representation of number, but letter comparison does not tap this same construct. Furthermore, children's magnitude (but not letter) comparison ability was a significant predictor of their concurrent arithmetic achievement (both untimed and speeded). A task that was initially included to act as a control measure, number identification, was also found to be a significant predictor of variance in children's arithmetic scores. This suggests that children who were better at comparing magnitudes (and possibly have more precise representations of number) and had a better knowledge of the symbolic number system achieved higher arithmetic scores.

Due to the lack of research investigating children's development on these comparison tasks longitudinally (rather than in cross-sectional designs), Chapter 3 explored children's development on the group presented comparison tasks. As expected, children's performance on all of the magnitude comparison tasks improved with age, reflecting a refinement in their internal representation of number (e.g. Halberda & Feigenson, 2008). Children also gained higher scores over time on the letter comparison task. The results also provided further support that these group administered pencil and paper comparison tasks are replicating the findings reported in the literature using individually presented computerised measures and therefore justify their use as potential predictors of arithmetic achievement in the present research.

Chapters 4 and 5 both aimed to investigate whether individual differences in magnitude comparison ability predicted individual differences in children's later arithmetic achievement. Chapter 4 used the WIAT Numerical Operations subtest as the arithmetic outcome measure; this test assesses children's basic understanding of number as well their calculation ability on a variety of numerical operations (addition, subtraction, multiplication and division), while Chapter 5 assessed children's calculation fluency on the numerical operations (addition, subtraction, and multiplication) separately. Children's prior arithmetic ability was controlled in the analysis so that *growth* in arithmetic skill could be examined. Moderate to strong correlations were observed between earlier and later arithmetic achievement (r = .53 to .75), which meant that there was potential for other tasks to predict growth. Children's performance on the number identification task was found to be a unique predictor of their growth on the WIAT arithmetic measure. In contrast children's magnitude comparison ability, while being related to later arithmetic achievement, was not found to predict growth in their scores.

An additional aim of Chapter 5 was to examine if the predictive relationship between number identification and WIAT arithmetic achievement was due to a possible overlap in the underlying skills being assessed, as the WIAT includes items such as writing Arabic numerals and identifying numbers from different symbols. It is also theoretically possible that different arithmetic operations depend on different precursor skills (Dehaene, 1992; Dehaene & Cohen, 1997). Only the speeded addition test was completed by all children at Time 1, so this measure was used as the control for children's prior arithmetic skill. Fluency at solving addition problems was a predictor of all later calculation fluency measures; therefore it appears that a good grounding in the ability to solve simple addition problems is also important for solving problems based on the other arithmetic operations. The results showed that number identification contributed to the growth of calculation fluency skill suggesting that the relationships observed between the same task and achievement on the WIAT arithmetic measure were not only due to a possible overlap between some items. The only exception to the finding of the importance of earlier number identification skill was subtraction ability at Time 2; in fact no other variable apart from prior addition skill explained individual differences in Time 2 subtraction scores. However it does not appear to be the case that number identification ability is unimportant for subtraction skill in general, as it was found to be a significant predictor of Time 3 subtraction achievement.

The contribution of magnitude comparison to the development of arithmetic fluency was significant for Time 2 addition (and was also marginal at Time 3). This could possibly reflect children's strategy use on the addition task; an efficient strategy that children could draw on to solve the problem would be to count on from the largest of the two addends in the problem and to do this they would first need to compare them (Butterworth et al., 2001). It has also been proposed in the literature that even when people are able to use a retrieval strategy to solve the problem arithmetic facts are stored as larger addend + smaller addend and comparison still remains as a stage in answering the sum (Butterworth et al., 2001). In contrast, magnitude comparison ability was not a significant predictor of children's later subtraction or multiplication ability. From these analyses it appears that, in general, the acuity of a child's internal representation of number at 6 years may only influence later addition achievement and not achievement on the other operations. This may explain why in the current research magnitude comparison ability was not a longitudinal predictor of children's arithmetic achievement assessed using the WIAT arithmetic measure as it tests all arithmetic operations (addition, subtraction, multiplication and division).

The inconsistencies in the relationships observed between the magnitude comparison tasks and later arithmetic achievement are in contrast to some results reported in the literature (e.g. symbolic: De Smedt et al., 2009; nonsymbolic: Libertus et al., 2013) but are in agreement with others (symbolic: Desoete et al., 2012; nonsymbolic: Sasanguie et al., 2013). A possible reason for magnitude comparison ability not being a significant predictor could be the inclusion of the control for the autoregressive effect, as many of the studies reported in the literature do not control for children's prior arithmetic skill. A second possibility is the methodology used; the comparison measures were group presented paper and pencil tasks rather than individually presented computerised measures. To explore the first possibility the analyses were repeated but this time removing the arithmetic control. To explore the second possibility a study was conducted with a subgroup of the sample (Chapter 6) and will be discussed later.

With the autoregressive effect removed, magnitude comparison ability still failed to make a significant contribution to children's later arithmetic achievement assessed using the WIAT but it is acknowledged that it showed a trend towards significance for Time 2. In contrast, magnitude comparison became a significant predictor of both Time 2 and 3 addition ability, and Time 3 multiplication (subtraction at this time point was also marginal). It was not found to be a predictor of children's subtraction ability at Time 2 or their achievement on the WIAT at Time 3. It could be suggested that the speeded arithmetic measures share method variance with the magnitude comparison tasks (i.e. both measures had a speeded element); however, while children's performance on the letter comparison task (which was matched to the digit comparison task) was found to be associated with later arithmetic achievement, this was not a significant predictor of variance in scores. Therefore, the predictive relationship found between magnitude comparison and calculation fluency must be due to more than both measures being speeded.

In summary, the results from the large group study suggest that magnitude comparison ability may contribute towards children's arithmetic development but it is no more useful as a predictive measure than assessing children's arithmetic achievement at that earlier point in time. Even in the same sample of children, the importance of magnitude comparison ability was inconsistent; therefore it is not surprising that the findings observed in the literature are inconclusive. The findings also suggest that the type of measure used to assess arithmetic may influence the results. In comparison to this, the number identification measure was a consistent predictor of children's later arithmetic achievement.

Chapter 6 therefore had multiple aims; the first was to investigate whether number identification would remain a significant predictor of later arithmetic achievement when performance on individually presented magnitude comparison measures was controlled. The second aim was to rule out the possibility that the inconsistent predictive relationships observed in the large group studies were due to the methodology used to present the magnitude comparison tasks (Study 2). Therefore the studies reported in this chapter used computerised magnitude (symbolic and nonsymbolic) comparison tasks more akin to those used by previous studies. Other questions raised from reviewing the literature regarded the reliability of nonsymbolic comparison tasks; recently-published studies have examined the test-retest reliability of these measures but more extensive research is still lacking, especially within the child population. Study 1 of this chapter therefore presented data from the completion of a nonsymbolic comparison task four times over an 18 month time period, with children completing the task twice at the second time point. Forth, an alternative and increasingly popular way to gain an estimate of the acuity of an individual's approximate number system (ANS), rather than just using overall accuracy or speed at comparing numerosities, is to calculate a Weber fraction using accuracy scores (see Halberda, et al., 2008). However, only one study has investigated the refinement of the ANS over development using the Weber fraction calculated for the same sample of children (Libertus et al., 2013), therefore Chapter 6 also explored this over an 18 month time period (Study 1).

In line with the concurrent literature, Study 1 found that children's overall accuracy on the nonsymbolic comparison task improved over time and the estimate of the Weber fraction also reduced in size, however, the change between Time 2 and 3, which was roughly 12 months, was not significant. It is possible that a slowing down in the refinement of the ANS takes place between 7 and 8 years old, although further research is needed over a longer development period to clarify this. At the second time point, children completed the nonsymbolic comparison task twice within the same testing session in order to gain an estimate of the test-retest reliability. Moderate significant correlations were observed between the two estimates of overall accuracy and the Weber fraction (r = .62 and .54 respectively). The associations observed between the different time points were weaker than those within the same session which questions the stability of the measure over time, and in general the Weber fraction was less reliable than overall accuracy scores. In addition, the internal consistency estimates were also low given the large number of items in the task. In summary, these findings severely question the use of the nonsymbolic comparison task as a longitudinal predictor.

Given the wide use of the nonsymbolic comparison task in existing studies that predict children's arithmetic development, children's overall accuracy scores were included in correlation and regression analyses alongside their performance on a symbolic comparison task. An additional reason was so that the analysis was comparable to that presented in the previous chapters. Controlling for age and general cognitive ability, neither of the magnitude comparison tasks were found to contribute unique variance to individual differences in later arithmetic achievement. In contrast, children's achievement on the number identification measure at Time 1 was again important for their later arithmetic achievement assessed almost one and two years later. To explore whether this relationship could be due to the number identification measure capturing variance related to counting skill, children's counting speed was included in place of magnitude comparison ability. The contribution of number identification to the development of arithmetic remained significant even with counting speed controlled.

To be comparable to the large group analysis, children's prior arithmetic skill was then controlled so that *growth* in arithmetic achievement could be explored. Number identification remained as a significant unique predictor of children's later calculation fluency (timed measures) but not of their development on the WIAT arithmetic measure (Time 2 showed a trend towards significance). It was therefore suggested that children's early knowledge of the symbolic number system may be even more important for calculation fluency skills where quick access to the quantity representations of numbers is required.

In summary, the studies and analysis presented in this thesis repeatedly found that the number identification task was a more consistent and better predictor of children's arithmetic development than the more frequently used magnitude comparison tasks that are thought to be an indicator of the acuity of an individual's internal representation of number (e.g. Halberda et al., 2008; Sasanguie et al., 2013). Number identification remains as a significant longitudinal predictor of growth in arithmetic fluency even when children's general cognitive ability, earlier arithmetic achievement and counting speed are controlled for.

So what does the number identification task assess? The number identification test constructed for these studies required children to match a verbally presented number to its symbolic form from multiple options with distractor items including errors that young children make when writing numbers (see Power & Dal Martello, 1990). The numbers included ranged from one digit (6) to three digits (427), so the task not only required children to know the symbolic number system but also to map (or transcode) between different representations of number (i.e. verbal to symbolic; triple code model, Dehaene, 1992) and assessed their understanding of place value. Consistent with previous literature using number writing tasks, it was found that when children made errors the most common one was to choose the symbol with additional zeros (syntactic error) rather than the option where the digits were correct but in the incorrect order (Power & Dal Martello, 1990; Seron & Fayol, 1994). Understanding of place value is important for arithmetic achievement, as even in simple addition when numbers are combined (or subtracted) the result needs to be integrated into one number and once this result is beyond a single digit, the rules of the Arabic number system needs to be applied (Moeller et al., 2011). Further research is needed to explore the different aspects of the type of number identification task used here and its relationship with arithmetic. A number mapping task where children have to transcribe, rather than identify, the symbolic number form may assist this. A number writing task was originally included in the current study, however, a large number of children were already consistently able to write two digit numbers correctly, which resulted in a lack of variance on this measure. This suggests that numbers outside this range (particularly three digit numbers) should be included in measures presented to this age group. Measures that assess children's understanding of place value also need to be developed. Ideally a symbolic number knowledge task that assesses the ability to identify, write, and even read numbers could be used as a screening procedure to identify children who may go on to develop weaker arithmetic knowledge than their peers. It would be advantageous to identify these children as early as possible, perhaps during the first year of formal schooling (reception in the UK); further research is therefore needed to develop measures like the one used in this study for use with younger children, and also longitudinal studies beginning earlier in development to investigate whether the findings hold.

In this thesis, the lack of consistent evidence for a predictive relationship between magnitude comparison ability and later arithmetic achievement suggests that the internal system for representing number does not form the foundation upon which formal arithmetic knowledge builds, in this age group at least. There are possible reasons why this may not have been found in the research presented here, first, the approximate number system may not be a precursor for more complex arithmetic skill to build upon. Second, magnitude comparison tasks may not be reliable and valid measures to estimate the acuity of internal magnitude representations. Third, while the sample of children involved in this research was young (on average 6 years old at Time 1), they had nonetheless already begun their formal schooling. It is possible that with development the importance of this system declines and other knowledge (for example of the symbolic number system) becomes more important. To clarify these points, further research is needed; more longitudinal studies are required beginning before children start formal schooling. Due to the inconsistent results reported in the wider literature as well as in this thesis, multiple nonsymbolic comparison tasks should be presented to the same group of children, which would allow for further investigation into the validity of these measures. However, we should also remain cautious when using these types of task due to the less than ideal estimates of reliability observed here and in the wider literature.

7.1. Future directions

7.1.1. Mapping between Number Codes

Linked with the number identification task is the idea that children's ability to map between number representations (verbal, symbolic and magnitude) might be important for children's arithmetic development. The symbolic comparison tasks used throughout this thesis involve an element of this as it is proposed that in order to compare digits they first need to be mapped onto magnitude representations (Moyer & Landauer, 1967). Using more direct mapping tasks (symbolic to nonsymbolic and nonsymbolic to symbolic), Mundy & Gilmore (2009) found that scores on the tasks (overall accuracy, symbolic to nonsymbolic and nonsymbolic to symbolic) were not related to arithmetic achievement. However, in regression analyses that only included performance on difficult mapping problems (ratio of .67 between target and distractor), mapping ability was found to be a significant predictor of concurrent arithmetic achievement after controlling for symbolic and nonsymbolic comparison accuracy. In line with this finding of the importance of mapping ability, Nys et al. (2013) reported that adults who had not received mathematics education, performed more poorly on mapping tasks than those that had received (at least some) mathematics education. The mapping tasks used by Mundy and Gilmore, did not specifically assess mapping ability between only two of the number representation codes; the symbolic numbers were presented with their verbal label so it is unclear whether the task was indeed tapping mapping ability between symbolic and nonsymbolic representations. Future studies could therefore more systematically investigate mapping ability between the three number representations in children who are in the early stages of their mathematics education.

7.1.2. Strategy Use

Another important point to consider is the strategies that children use to solve arithmetic problems. Even when a child is presented with a simple addition sum they can draw upon a range of strategies that they may have been taught or created themselves in order to solve it. Evidence shows that with development (and practice), the strategies that children apply become more efficient (e.g. Farrington-Flint et al., 2009; Lemaire & Siegler, 1995; Siegler, 1987), and it has been found that efficiency at solving simple arithmetic problems predicted growth in mathematics ability (e.g. Hecht et al., 2001). However, there is considerable heterogeneity and flexibility of strategy use both within and between children (see Siegler, 1987; 1999). Children’s strategy choice as well as their flexibility in applying strategies may be important for not only solving simple calculation problems but also when it comes to solving more complex multidigit problems. This may link with domain-general skills, i.e. working memory and attention (discussed below). Future research is needed exploring early strategy use and its relation to later arithmetic and mathematics achievement.

7.1.3. Domain-general Predictors

This thesis has focused on domain-specific predictors of arithmetic. In the literature review it was highlighted that domain-general predictors are important too, for example aspects of working memory have been found to be predictors of children’s arithmetic achievement (see Friso-van den Bos et al., 2013). It was also noted that some of the studies investigating the relationship between magnitude comparison ability and arithmetic explored this whilst controlling for domain-general skills. The findings suggest that the relationship between the approximate number system (assessed using magnitude comparison tasks) and arithmetic/mathematics achievement is not attributable to domain-general skills.

A less well researched area is whether inhibitory control skills play a part in this relationship. In two recent studies Fuhs and McNeil (2013) and Gilmore et al. (2013) found that the relationship between children’s accuracy on nonsymbolic comparison tasks and arithmetic was driven by performance on incongruent trials, this is where the surface area of the stimuli conflicts with numerosity (i.e. as numerosity increases the surface area decreases), which requires children to inhibit a response based on surface area. Both studies also reported that in regression analyses when a measure of inhibitory control skill was entered before performance on the nonsymbolic comparison task it was no longer a significant predictor of arithmetic achievement. The nonsymbolic comparison tasks used throughout this thesis did not include trials where children had to inhibit a response based on surface area, which could possibly contribute to the lack of a predictive relationship of magnitude comparison to later arithmetic achievement. However, some of the comparisons in the group presented tasks involved choosing the box that had the *smaller* squares (total surface area was equated meaning that the larger numerosity included smaller squares). Children therefore had to inhibit the response of choosing larger squares in favour of choosing the larger numerosity as this would have lead to an incorrect response. In addition, symbolic comparison tasks do not involve inhibiting a response based on other characteristics and it was found that these symbolic and nonsymbolic comparison tasks loaded on the same factor so this cannot fully explain the lack of a longitudinal predictive relationship found (magnitude comparison was a significant correlate with arithmetic in the large group studies but not always a significant unique predictor). Future studies should therefore include measures of inhibitory control when investigating the relationship between magnitude comparison performance and arithmetic achievement.

Another domain-general skill that has been found to be important for children's arithmetic development is attention (e.g. Fuchs et al., 2005; Fuchs et al., 2006). It is also entirely possible that factors such as attention could provide alternative explanations for some aspects of the data reported here. For example, the computerised nonsymbolic comparison task presented in Chapter 6 required children to concentrate for 120 experimental trials; the task also included some very difficult comparison items so they had to stay motivated to complete the task. Impulsivity may also play a role as some children may have just been pressing a response without actually comparing the items in order to complete the task quickly. Taken together these factors could explain the less than ideal measures of reliability obtained for this measure. On a positive note, many of the measures administered during the group testing phases of the research presented here were short and involved a timed element in order to sustain children's attention and to keep them motivated. It is acknowledged that some children may have shown poor concentration and impulsivity on the group comparison measures, for example ticking down one side of the page, but on reflection of scoring these tasks this only occurred for a small number of children even at the first time point.

Future research should therefore include measures of executive functions, inattention and impulsivity and investigate the unique contribution that magnitude comparison ability makes to children's arithmetic development. The importance of some factors may also change with development and experience, for example nonsymbolic magnitude comparison ability may well be important when children are at the early stage of their numeracy development but as they begin to learn the symbolic number system a measure of this knowledge (i.e. number identification, symbolic comparison ability) may be more useful. As children then begin to learn arithmetic, working memory and attention may play a greater role. Therefore it would be advantageous to explore the unique contributions of potential predictors alongside each other rather than looking at specific abilities in isolation (e.g. Cirino, 2011), and whether some skills act as mediators, for example working memory (e.g. Geary, Hoard, Byrd-Craven, Nugent & Numtee, 2007).

7.2. Conclusion

In summary, the research presented in this thesis consistently found that children's ability to match a verbally presented number to its symbolic form (number identification) was important for their arithmetic development. In contrast, the relationship between children's performance on magnitude comparison tasks and arithmetic was found to be inconsistent; it differed across the type of measure used to assess children's arithmetic achievement and the method used to assess comparison ability. It cannot be concluded that a child's internal representation of number is not important for their arithmetic development at all as it is entirely possible that this relationship is important much earlier in development or dependent on the methodology used. However, the central message to take from this research is that in this age group of children (5 to 8 years old) it is imperative that they are familiar with the symbolic number system, including mapping between verbal and symbolic number forms, understand place value and know the rules for transcribing multidigit numbers, in order to thrive at arithmetic. This task was a simple group presented paper and pencil assessment that could be used to identify children who may go on to develop weaker arithmetic ability than their peers. Future research should therefore explore this further and in turn develop intervention methods to help improve these skills in children who are struggling with arithmetic.

Appendix 1.

WIAT Numerical operations test presented at Time 1, 2 and 3

Items presented at all time points -1 to 15

1. Identify numbers from presented numbers and letters, had to circle 5 items
2. Identify the number 6 from presented numbers, had to circle 3 items
3. Fill in the missing number 7 on a number line of 1 to 10
4. Write the numeral 3
5. Write the numeral 10
6. Count 8 dots and write how many were counted
7. 3 + 3 =
8. 8 + 5 =
9. 4 – 2=
10. 2 + 3 + 1 + 4 =
11. 10 – 6 (presented as a column sum)
12. 41 + 14 =
13. 68 – 43 (presented as a column sum)
14. 8 * 5 =
15. 37 + 54 (presented as a column sum)

Items added at Time 2

16. 16 ÷2 =
17. 120 – 15 =
18. 7 * 6 =

Items added at Time 3

19. 24 x 5 (presented as a column sum)
20. 80 – 56 (presented as a column sum)
21. 698 + 426 (presented as a column sum)
22. 69 ÷ 3 =
23. 57 + 32 + 94 + 48 (presented as a column sum)
24. 0.2 + 0.8 =
25. $\frac{7}{8} - \frac{3}{8} =$ _____
26. £5.31 – 2.47 (presented as a column sum)
27. 705 – 489 (presented as a column sum)

Appendix 2.

One minute addition test presented at all time points

Name ________________

Date ___________

Adding up

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

I

Appendix 3.

One minute subtraction test presented at all time points

Name ____________________   Date ______________

Taking away

2 – 1 =	8 – 5 =
5 – 1 =	9 – 5 =
3 – 2 =	10 – 4 =
5 – 3 =	9 – 4 =
6 – 2 =	10 – 3 =
2 – 2 =	11 – 2 =
6 – 4 =	10 – 6 =
7 – 2 =	12 – 3 =
6 – 1 =	12 – 6 =
7 – 3 =	11 – 5 =
8 – 2 =	13 – 3 =
7 – 5 =	12 – 9 =
8 – 3 =	14 – 6 =
7 – 4 =	17 – 8 =
9 – 3 =	16 – 9 =

Appendix 4.

Number identification task

Name_______________________

Which is the right number?

a	8	6	3	9	
b	1	41	4	14	
c	82	28	208	8	20
d	502	5	25	50	52
e	76	17	6	706	67
f	25	235	20035	23	253
g	13	10063	136	15	163
h	472	427	47	42	40027

Appendix 5.

Presentation order of all comparison tasks at Time 1

Booklet A

1. Digit mixed
2. Nonsymbolic far SA (surface area matched)
3. Nonsymbolic 5:6 SS (same size)
4. Letter close
5. Nonsymbolic 3:4 SS (same size)
6. Nonsymbolic mixed SA (surface area matched)
7. Digit far
8. Nonsymbolic 7:8 SA (surface area matched)
9. Nonsymbolic close SS (same size)

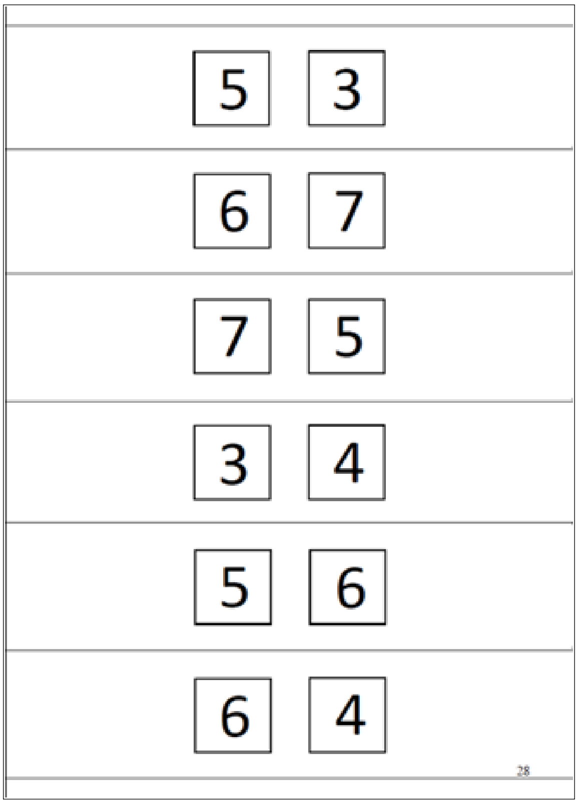
Booklet B

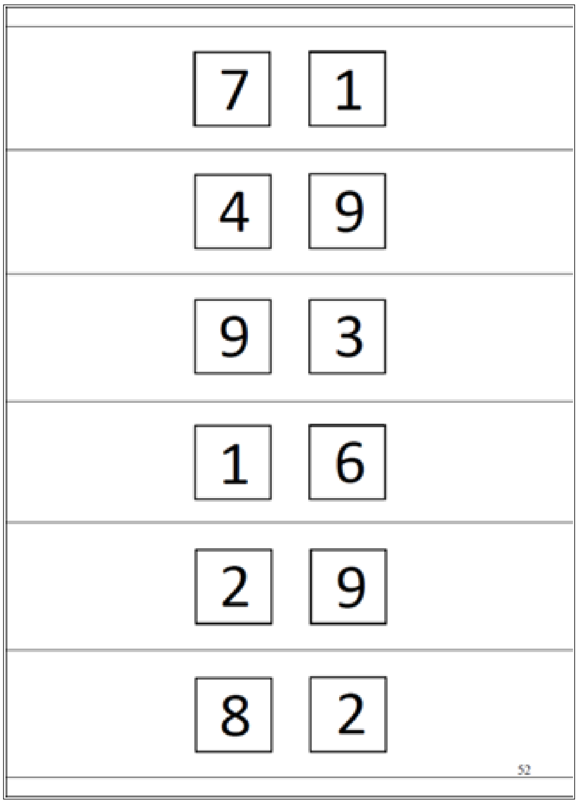
1. Letter far
2. Nonsymbolic close SA (surface area matched)
3. Nonsymbolic 7:8 SS (same size)
4. Digit close
5. Nonsymbolic far SS (same size)
6. Nonsymbolic 5:6 SA (surface area matched)
7. Letter mixed
8. Nonsymbolic mixed SS (same size)
9. Nonsymbolic 3:4 SA (surface area matched)

Appendix 6.

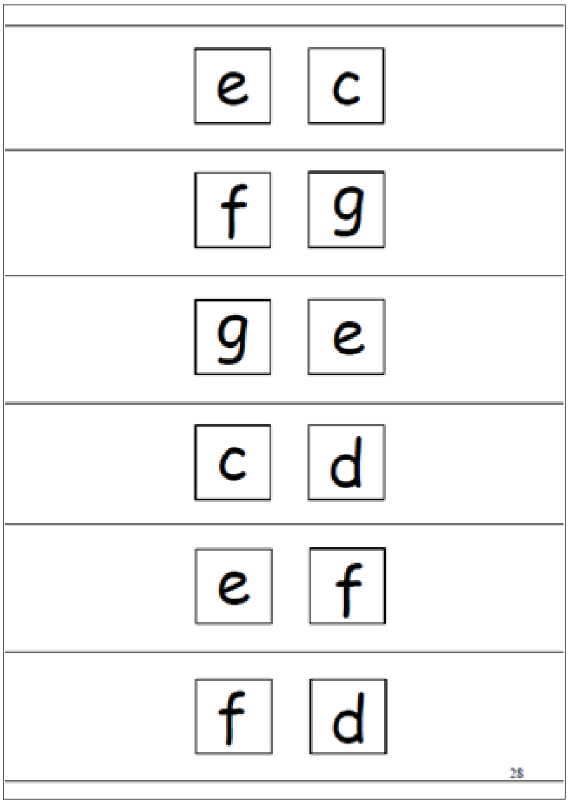
Example of the comparison tasks

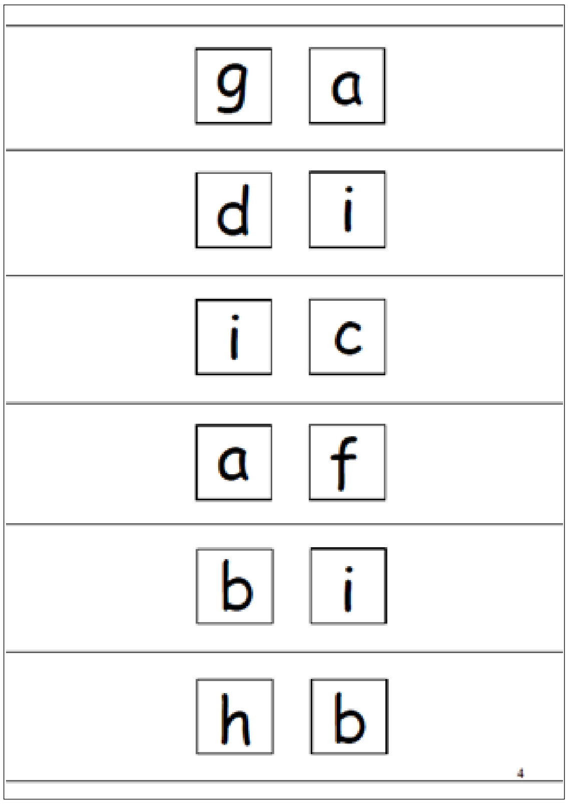
Digit close (left) and digit far (right)



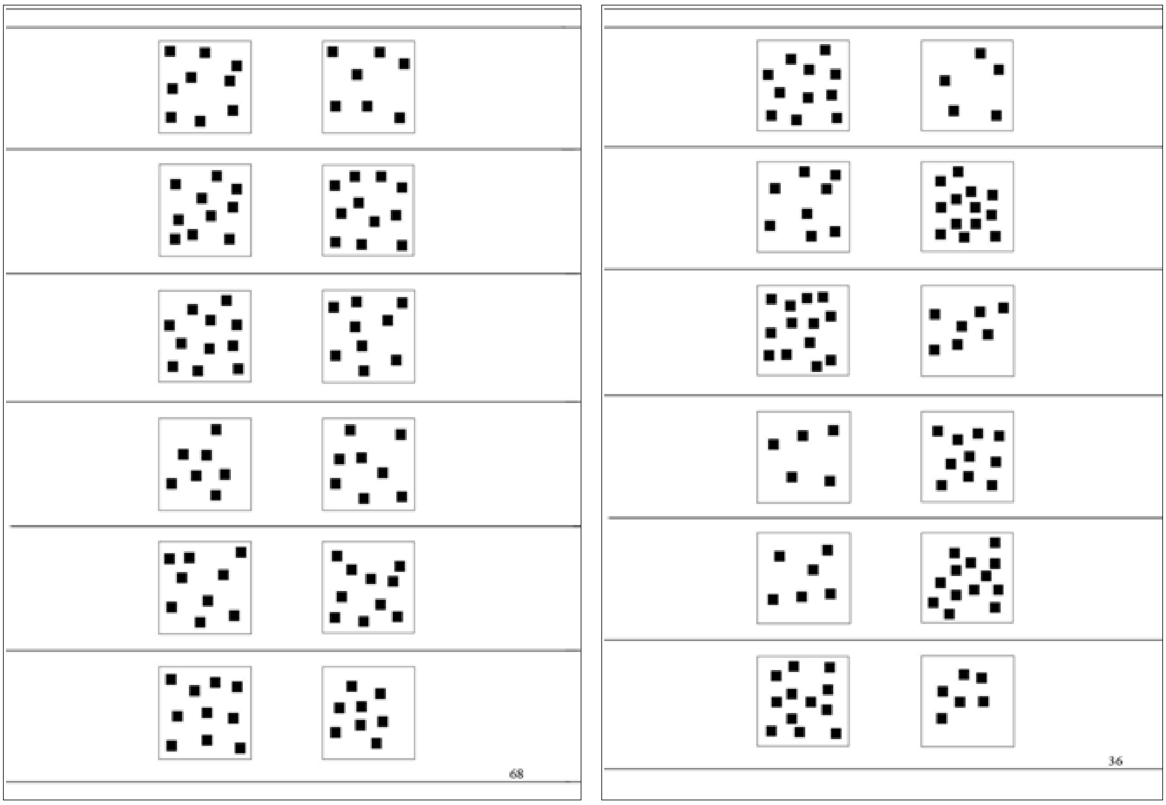


Letter close (left) and letter far (right)

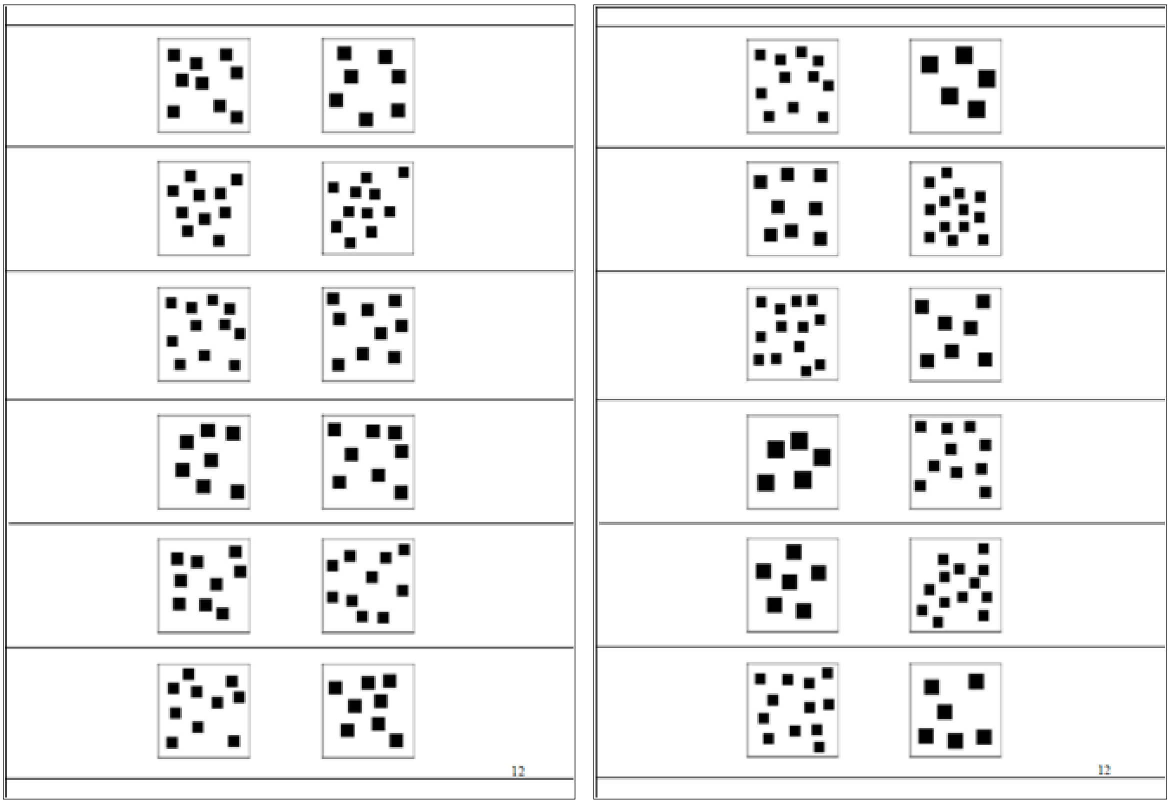




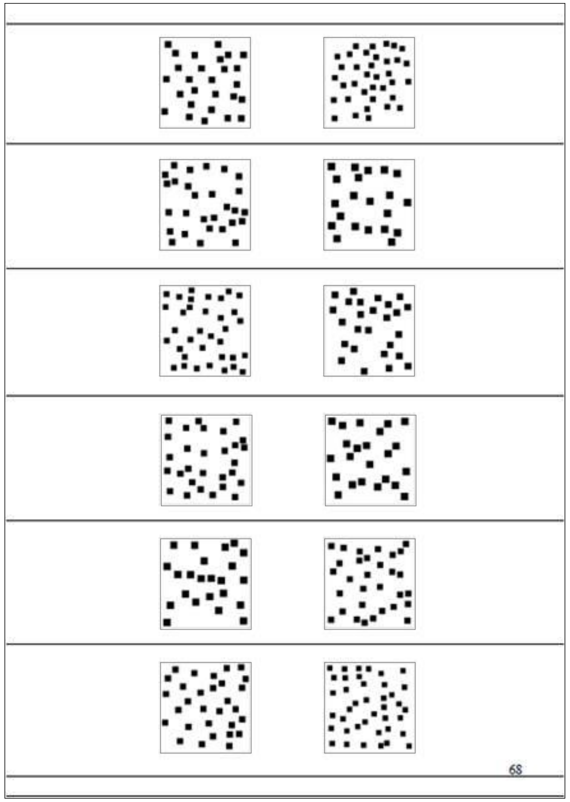
Nonsymbolic close same size stimuli (left) and nonsymbolic far same size stimuli (right)

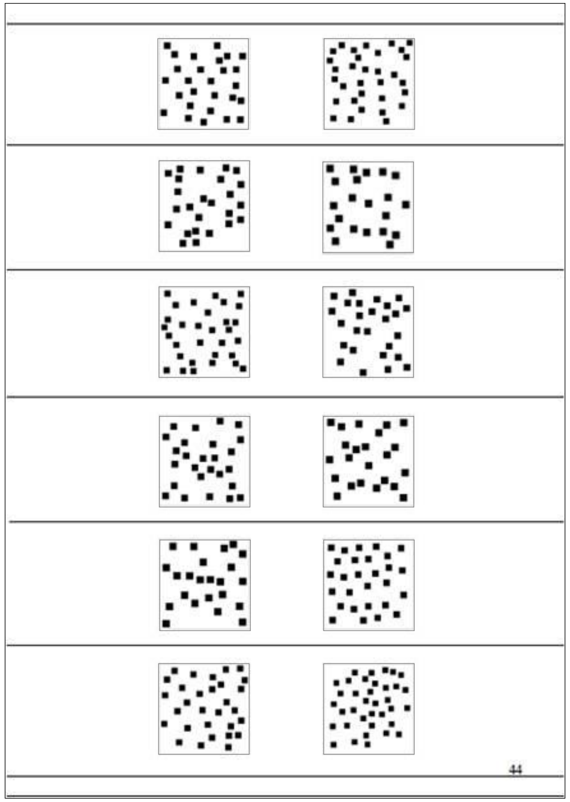


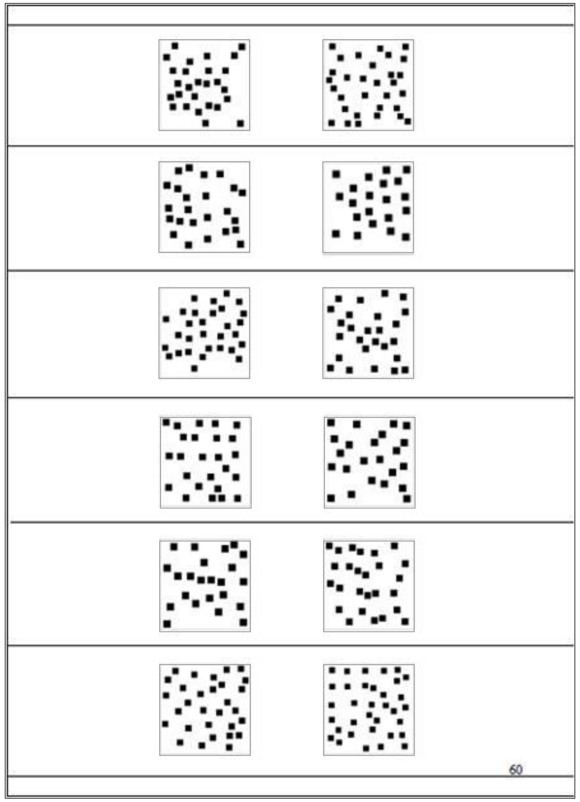
Nonsymbolic close surface area matched stimuli (left) and nonsymbolic far surface area matched stimuli (right)



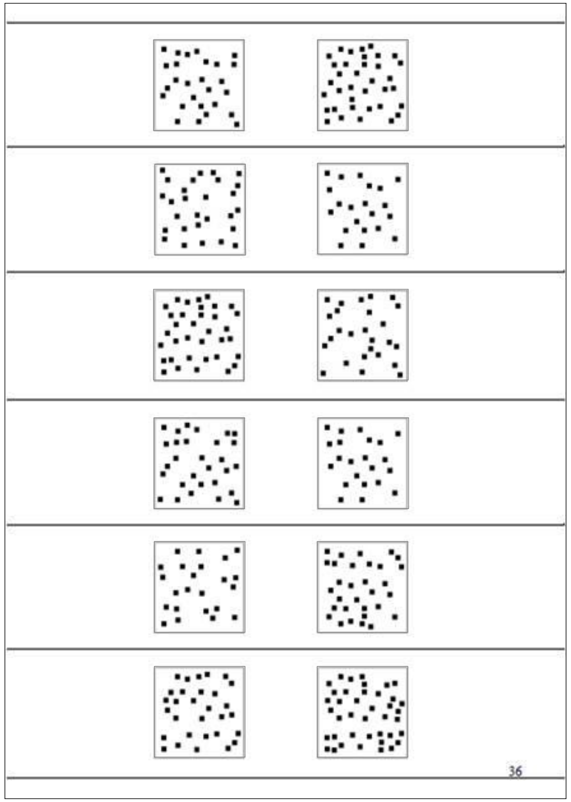
Nonsymbolic ratio manipulation: 3:4, 5:6, 7:8 surface area matched stimuli

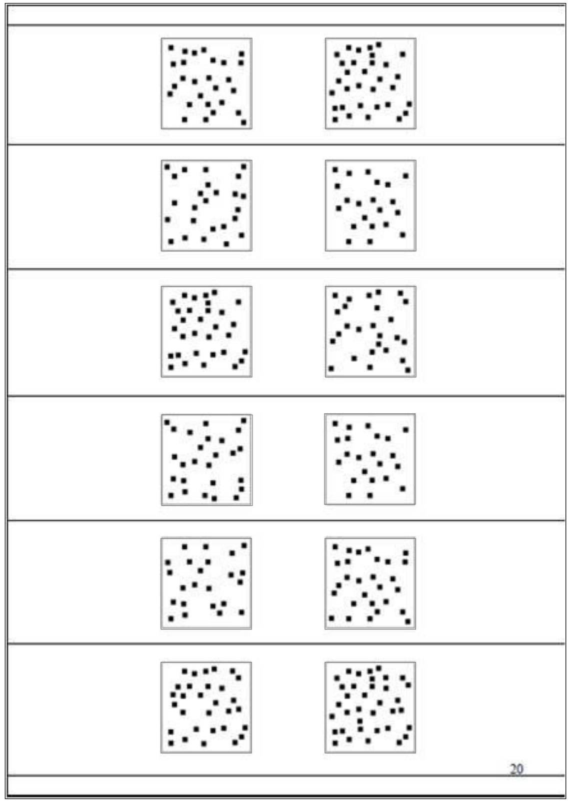


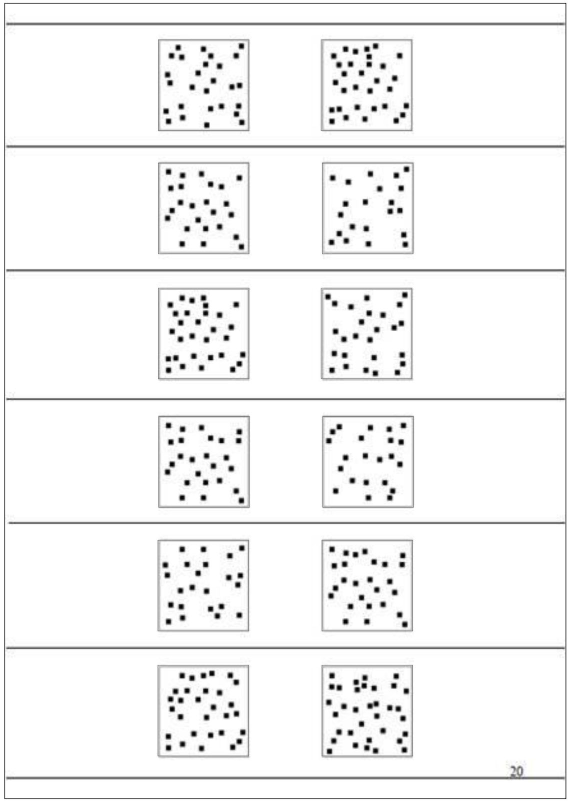




Nonsymbolic ratio manipulation: 3:4, 5:6, 7:8 same size stimuli







Appendix 7.
Construction of nonsymbolic stimuli

For the investigation of the effect of distance (numerosity range = 5 to 13)

For the surface area matched comparison items each stimulus had a total black area of 6000 paintbrush dots (this refers to the smallest paintbrush available on Paint). Six thousand was divided by the numerosity of the stimulus (5 to 13) to give the amount of paintbrush dots for each square within that stimulus. The square root was taken to give the width and length of each square. Each square within a numerosity was made on 800% zoom. Each stimulus item was then constructed on 100% zoom within a $5cm^2$ border; each square was placed randomly within this border. Four versions of each numerosity were constructed with different placing of the individual squares. The stimuli were then reduced to $2.5cm^2$ in size when they were presented to children in the booklets. For the same size comparison items the squares used for the stimuli were the same size as those created for the presentation of 13 squares in the surface area matched condition. Apart from the size of each square being the same in each stimulus they were created and presented in the same way as the surface area controlled stimuli.

For the investigation of the effect of distance (numerosity range = 20 to 40)

For the surface area matched each stimulus had a total black area of 1,000,000 paintbrush dots (again this refers to the smallest paintbrush available on Paint). One million was divided by the numerosity of the stimulus (20 to 40) to give the amount of paintbrush dots for each square within that stimulus. The square root was calculated to give the width and length of each square (in pixels) which were created using ArcSoft PhotoStudio 5.5. This single square was transferred to Paint and copied to give the desired numerosity. Each stimulus was constructed in a 2750 x 2750 pixel sheet ($72.75cm^2$). Again this was reduced down to $2.5cm^2$ in size when presented to the children. The squares used for the same size stimuli were the same size as those created for the presentation of the numerosity 45 in the SA condition. Apart from the size of each square being the same in each stimulus they were created and presented in the same way as the surface area controlled stimuli.

Appendix 8.

Presentation order of measures at Time 1 for each school

1	2	3a	3b	3c	4a	4b
Comparison A	Comparison A	Comparison A	Comparison A	Comparison A	Comparison A	Comparison A
Number writing	Ravens	Ravens	Ravens	Ravens	Ravens	Ravens
Number ID	1 minute add	1 minute add	1 minute add	1 minute add	1 minute add	1 minute add
Ravens	BPVS	Number writing	Number writing	Number writing	(1 minute sub)	(1 minute sub)
1 minute add	Comparison B	Number ID	Number ID	Number ID	BPVS	BPVS
BPVS	WIAT arithmetic	BPVS	BPVS	BPVS	WIAT arithmetic	WIAT arithmetic
Comparison B	Spelling	WIAT arithmetic	WIAT arithmetic	WIAT arithmetic	Number writing	Number writing
WIAT arithmetic	BPVS	Comparison B	Comparison B	Comparison B	Number ID	Comparison B
BPVS	Number writing	BPVS	BPVS	Spelling	Comparison B	BPVS
Spelling	Number ID	Spelling	Spelling	1 minute sub	BPVS	Spelling
	1 minute sub		1 minute sub		Spelling	Number ID

Note: With schools 3 and 4 there were more than one class of children so they were tested as different groups. Children at school 4 were tested according to their numeracy grouping at school (e.g. higher achievers and lower achievers) on the first session, for sessions 2 and 3 they were tested with the other children in their class. Both groups were administered comparison A and Ravens in the first session but only one group also received 1 minute addition. Therefore in session 2 when the children were tested according to their class groupings children who had already completed 1 minute add completed 1 minute sub instead.

Appendix 9.

Standardized coefficients between the variables in Figure 2.13: Path model at Time 1 predicting WIAT Numerical Operations

	Estimate	Two tailed p-value
Age with		
Raven	0.177	0.046
BPVS	0.312	0.001
Comparison judgement	0.291	0.000
Letter comparison	0.230	0.006
Number ID	0.202	0.023
Raven with		
BPVS	0.236	0.036
Comparison judgement	0.451	0.000
Letter comparison	0.537	0.000
Number ID	0.394	0.000
BPVS with		
Comparison judgement	0.311	0.002
Letter comparison	0.266	0.014
Number ID	0.623	0.000
Comparison judgement with		
Letter comparison	0.649	0.000
Number ID	0.445	0.000
Letter comparison with		
Number ID	0.385	0.000

Appendix 10.

Standardized coefficients between the variables in Figure 2.15: Path model at Time 1 predicting one minute addition

	Estimate	Two tailed *p*-value
Age with		
Raven	0.176	0.047
BPVS	0.317	0.000
Comparison judgement	0.291	0.000
Letter comparison	0.228	0.009
Number ID	0.203	0.022
Raven with		
BPVS	0.236	0.037
Comparison judgement	0.451	0.000
Letter comparison	0.550	0.000
Number ID	0.396	0.000
BPVS with		
Comparison judgement	0.314	0.001
Letter comparison	0.273	0.014
Number ID	0.632	0.000
Comparison judgement with		
Letter comparison	0.669	0.000
Number ID	0.442	0.000
Letter comparison with		
Number ID	0.397	0.000

Appendix 11.

Presentation order of measures at Time 2 for all schools

Session 1:

1. WIAT Numerical Operations
2. Spelling
3. One minute addition
4. One minute subtraction
5. One minute multiplication
6. Number writing
7. Number identification

Session 2:

8. Bespoke untimed arithmetic test
9. Comparison tasks

If time allowed:

10. Four bespoke one minute addition tests

Appendix 12.

Presentation order of measures at Time 3 for all schools

Session 1:

1. One minute addition
2. One minute subtraction
3. One minute multiplication
4. WIAT: Numerical Operations
5. Number copying
6. Addition without a carryover (totals less than 20)
7. Addition with a carryover (totals less than 20)
8. Spelling (where time allowed)

Session 2:

9. Spelling
10. Comparison tasks
11. (Remaining spelling items if needed)

Appendix 13.

Order of comparison tasks at Time 2 and Time 3

1. Practice- Digit mixed
2. Practice - Nonsymbolic mixed SS (same size)
3. Letter close
4. Digit far
5. Nonsymbolic 7:8 SA (surface area matched)
6. Nonsymbolic close SS (same size)
7. Letter far
8. Digit close
9. Nonsymbolic far SS (same size)
10. Nonsymbolic 56 SA (surface area matched)
11. Letters additional task
12. Nonsymbolic 3:4 SA (surface area matched)

At Time 3 a speed of processing task was added, this was presented after the letter far comparison task.

Appendix 14.

Standardized coefficients between the variables in Figure 4.1: Path model predicting Time 2 WIAT Numerical Operations

	Estimate	Two tailed *p*-value
Age with		
Raven	0.180	0.041
BPVS	0.314	0.001
WIAT	0.185	0.030
Comparison judgement	0.290	0.000
Letter comparison	0.231	0.007
Number ID	0.208	0.018
Raven with		
BPVS	0.240	0.033
WIAT	0.451	0.000
Comparison judgement	0.453	0.000
Letter comparison	0.544	0.000
Number ID	0.402	0.000
BPVS with		
WIAT	0.522	0.000
Comparison judgement	0.314	0.001
Letter comparison	0.274	0.012
Number ID	0.625	0.000
WIAT with		
Comparison judgement	0.537	0.000
Letter comparison	0.442	0.000
Number ID	0.660	0.000
Comparison judgement with		
Letter comparison	0.660	0.000
Number ID	0.448	0.000
Letter comparison with		
Number ID	0.392	0.000

Appendix 15.

Standardized coefficients between the variables in Figure 4.2: Path model predicting Time 2 WIAT Numerical Operations (without the autoregressor)

	Estimate	Two tailed *p*-value
Age with		
Raven	0.180	0.042
BPVS	0.320	0.000
Magnitude comparison	0.291	0.000
Letter comparison	0.232	0.006
Number ID	0.215	0.014
Raven with		
BPVS	0.235	0.037
Magnitude comparison	0.452	0.000
Letter comparison	0.543	0.000
Number ID	0.397	0.000
BPVS with		
Magnitude comparison	0.321	0.001
Letter comparison	0.272	0.012
Number ID	0.634	0.000
Magnitude comparison with		
Letter comparison	0.659	0.000
Number ID	0.455	0.000
Letter comparison with		
Number ID	0.389	0.000

Appendix 16.

Standardized coefficients between the variables in Figure 4.3: Path model predicting Time 3 WIAT

	Estimate	Two tailed *p*-value
Age with		
Raven	0.180	0.042
BPVS	0.313	0.001
WIAT	0.187	0.028
Magnitude comparison	0.291	0.000
Letter comparison	0.229	0.009
Number ID	0.202	0.022
Raven with		
BPVS	0.242	0.032
WIAT	0.454	0.000
Magnitude comparison	0.453	0.000
Letter comparison	0.550	0.000
Number ID	0.399	0.000
BPVS with		
WIAT	0.521	0.000
Magnitude comparison	0.310	0.002
Letter comparison	0.281	0.011
Number ID	0.630	0.000
WIAT with		
Magnitude comparison	0.538	0.000
Letter comparison	0.456	0.000
Number ID	0.655	0.000
Magnitude comparison with		
Letter comparison	0.673	0.000
Number ID	0.439	0.000
Letter comparison with		
Number ID	0.405	0.000

Appendix 17.

Standardized coefficients between the variables in Figure 4.4: Path model predicting Time 3 WIAT Numerical Operations (without the autoregressor)

	Estimate	Two tailed *p*-value
Age with		
Raven	0.179	0.042
BPVS	0.318	0.000
Comparison judgement	0.292	0.000
Letter comparison	0.230	0.009
Number ID	0.206	0.019
Raven with		
BPVS	0.236	0.036
Comparison judgement	0.450	0.000
Letter comparison	0.550	0.000
Number ID	0.396	0.000
BPVS with		
Comparison judgement	0.311	0.002
Letter comparison	0.277	0.013
Number ID	0.637	0.000
Comparison judgement with		
Letter comparison	0.674	0.000
Number ID	0.440	0.000
Letter comparison with		
Number ID	0.402	0.000

Appendix 18.

One minute multiplication test presented at Time 3

Times Tables

2 x 2 =	6 x 5 =
10 x 4 =	5 x 5 =
8 x 10 =	5 x 9 =
7 x 10 =	6 x 2 =
10 x 2 =	3 x 5 =
2 x 9 =	10 x 6 =
2 x 3 =	10 x 9 =
7 x 2 =	2 x 1 =
5 x 2 =	8 x 5 =
10 x 10 =	2 x 8 =
1 x 10 =	5 x 1 =
2 x 4 =	5 x 7 =
10 x 5 =	4 x 3 =
10 x 3 =	8 x 3 =
5 x 4 =	1 x 3 =

Appendix 19.

Standardized coefficients between the variables in Figure 5.1: Path model predicting Time 2 one minute addition and Figure 5.2: Path model predicting Time 2 one minute subtraction

	One minute addition		One minute subtraction	
	Estimate	Two tailed *p*-value	Estimate	Two tailed *p*-value
Age with				
Raven	0.178	0.044	0.180	0.041
BPVS	0.316	0.000	0.318	0.000
1 minute addition	0.294	0.000	0.300	0.000
Magnitude comparison	0.290	0.000	0.291	0.000
Letter comparison	0.227	0.009	0.221	0.014
Number ID	0.204	0.021	0.208	0.018
Raven with				
BPVS	0.236	0.037	0.239	0.034
1 minute addition	0.416	0.000	0.414	0.000
Magnitude comparison	0.452	0.000	0.453	0.000
Letter comparison	0.548	0.000	0.557	0.000
Number ID	0.392	0.000	0.403	0.000
BPVS with				
1 minute addition	0.250	0.007	0.258	0.006
Magnitude comparison	0.314	0.001	0.319	0.001
Letter comparison	0.274	0.012	0.283	0.013
Number ID	0.626	0.000	0.636	0.000
1 minute addition with				
Magnitude comparison	0.540	0.000	0.537	0.000
Letter comparison	0.483	0.000	0.527	0.000
Number ID	0.467	0.000	0.470	0.000
Magnitude comparison with				
Letter comparison	0.667	0.000	0.681	0.000
Number ID	0.443	0.000	0.455	0.000
Letter comparison with				
Number ID	0.395	0.000	0.423	0.000

Appendix 20.

Standardized coefficients between the variables in Figure 5.3: Path model predicting Time 3 one minute addition, Figure 5.4: Path model predicting Time 3 one minute subtraction, Figure 5.5: Path model predicting Time 3 one minute multiplication

	One minute addition		One minute subtraction		One minute multiplication	
	Estimate	Two tailed *p*-value	Estimate	Two tailed *p*-value	Estimate	Two tailed *p*-value
Age with						
Raven	0.177	0.045	0.178	0.044	0.177	0.045
BPVS	0.314	0.001	0.317	0.000	0.312	0.001
1 minute addition	0.294	0.000	0.296	0.000	0.291	0.000
Magnitude comparison	0.290	0.000	0.291	0.000	0.290	0.000
Letter comparison	0.228	0.009	0.226	0.011	0.228	0.009
Number ID	0.198	0.025	0.204	0.021	0.199	0.024
Raven with						
BPVS	0.239	0.034	0.240	0.034	0.240	0.033
1 minute addition	0.419	0.000	0.416	0.000	0.416	0.000
Magnitude comparison	0.453	0.000	0.451	0.000	0.453	0.000
Letter comparison	0.550	0.000	0.552	0.000	0.550	0.000
Number ID	0.396	0.000	0.396	0.000	0.398	0.000
BPVS with						
1 minute addition	0.261	0.005	0.260	0.005	0.263	0.005
Magnitude comparison	0.315	0.001	0.316	0.001	0.324	0.001
Letter comparison	0.277	0.013	0.279	0.013	0.273	0.014
Number ID	0.627	0.000	0.638	0.000	0.637	0.000
1 minute addition with						
Magnitude comparison	0.544	0.000	0.539	0.000	0.545	0.000
Letter comparison	0.496	0.000	0.513	0.000	0.491	0.000
Number ID	0.478	0.000	0.470	0.000	0.475	0.000
Magnitude comparison with						
Letter comparison	0.669	0.000	0.676	0.000	0.669	0.000
Number ID	0.445	0.000	0.443	0.000	0.455	0.000
Letter comparison with						
Number ID	0.392	0.000	0.406	0.000	0.393	0.000

Appendix 21.

Standardized coefficients between the variables in Figure 5.6: Path model predicting Time 2 one minute addition (without the arithmetic control), Figure 5.7: Path model predicting Time 2 one subtraction (without the arithmetic control)

	One minute addition		One minute subtraction	
	Estimate	Two tailed *p*-value	Estimate	Two tailed *p*-value
Age with				
Raven	0.178	0.045	0.180	0.042
BPVS	0.315	0.001	0.317	0.000
Magnitude comparison	0.289	0.000	0.291	0.000
Letter comparison	0.229	0.007	0.225	0.011
Number ID	0.203	0.021	0.209	0.017
Raven with				
BPVS	0.236	0.036	0.239	0.034
Magnitude comparison	0.451	0.000	0.452	0.000
Letter comparison	0.541	0.000	0.556	0.000
Number ID	0.393	0.000	0.403	0.000
BPVS with				
Magnitude comparison	0.314	0.001	0.319	0.001
Letter comparison	0.269	0.014	0.281	0.012
Number ID	0.625	0.000	0.639	0.000
Magnitude comparison with				
Letter comparison	0.655	0.000	0.676	0.000
Number ID	0.443	0.000	0.461	0.000
Letter comparison with				
Number ID	0.386	0.000	0.422	0.000

Appendix 22.

Standardized coefficients between the variables in Figure 5.8: Path model predicting Time 3 one minute addition (without the arithmetic control), Figure 5.9: Path model predicting Time 3 one minute subtraction (without the arithmetic control), Figure 5.10: Path model predicting Time 3 one minute multiplication (without the arithmetic control)

	One minute addition		One minute subtraction		One minute multiplication	
	Estimate	Two tailed *p*-value	Estimate	Two tailed *p*-value	Estimate	Two tailed *p*-value
Age with						
Raven	0.177	0.045	0.178	0.044	0.177	0.045
BPVS	0.313	0.001	0.317	0.000	0.310	0.001
Magnitude comparison	0.291	0.000	0.292	0.000	0.291	0.000
Letter comparison	0.231	0.006	0.230	0.008	0.232	0.006
Number ID	0.197	0.025	0.204	0.021	0.199	0.024
Raven with						
BPVS	0.240	0.033	0.240	0.034	0.240	0.033
Magnitude comparison	0.452	0.000	0.450	0.000	0.452	0.000
Letter comparison	0.535	0.000	0.547	0.000	0.540	0.000
Number ID	0.397	0.000	0.395	0.000	0.398	0.000
BPVS with						
Magnitude comparison	0.315	0.001	0.316	0.001	0.322	0.001
Letter comparison	0.267	0.014	0.275	0.013	0.265	0.015
Number ID	0.628	0.000	0.639	0.000	0.639	0.000
Magnitude comparison with						
Letter comparison	0.647	0.000	0.667	0.000	0.652	0.000
Number ID	0.446	0.000	0.445	0.000	0.460	0.000
Letter comparison with						
Number ID	0.373	0.000	0.398	0.000	0.381	0.000

References

Adams, J. W., & Hitch, G. J. (1997). Working memory and children's mental addition. *Journal of Experimental Child Psychology, 67*(1), 21-38. doi: doi.org/10.1006/jecp.1997.2397

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*(1), 20-29. doi: http://dx.doi.org/10.1016/j.jecp.2009.11.003

Amelang, M., & Zielinski, W. (1997). *Psychologische diagnostik und intervention* (2nd ed.). Berlin: Springer.

Ansari, D., & Karmiloff-Smith, A. (2002). Atypical trajectories of number development: A neuroconstructivist perspective. *Trends in Cognitive Sciences, 6*(12), 511-516. doi: 10.1016/S1364-6613(02)02040-5

Aunola, K., Leskinen, E., Lerkkanen, M.-K., & Nurmi, J.-E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology, 96*(4), 699-713. doi: 10.1037/0022-0663.96.4.699

Baddeley, A. D. (1997). *Human memory: Theory and Practice (Revised edition)*. Hove, UK: Psychology Press.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423. doi: http://dx.doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-90). New York: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). Cambridge: Cambridge University Press.

Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition, 98*(3), 199-222. doi: 10.1016/j.cognition.2004.09.011

Barth, H., La Mont, K., Lipton, J., & Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proceedings of the National Academy of Sciences, 102*(39), 14116-14121. doi: 10.1073/pnas.0505512102

Baudonck, M., Debusschere, A., Dewulf, B., Samyn, F., Vercaemst, V., & Desoete, A. (2006). *De Kortrijkse Rekentest Revision (KRT-R). [The Kortrijk Arithmetic Test-Revision KRT-R].* Kortrijk, Belgium: CAR Overleie.

Berch, D. B., Foley, E. J., Hill, R. J., & Ryan, P. M. (1999). Extracting parity and magnitude from Arabic numerals: Developmental changes in number processing and mental representation. *Journal of Experimental Child Psychology, 74*(4), 286-308. doi: 10.1006/jecp.1999.2518

Bonny, J. W., & Lourenco, S. F. (2013). The approximate number system and its relation to early math achievement: Evidence from the preschool years. *Journal of Experimental Child Psychology, 114*(3), 375-388. doi: 10.1016/j.jecp.2012.09.015

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230-258. doi: 10.1177/0049124192021002005

Bugden, S., & Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition, 118*(1), 32-44. doi: 10.1016/j.cognition.2010.09.005

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*(3), 205-228. doi: 10.1080/87565640801982312

Butterworth, B., Zorzi, M., Girelli, L., & Jonckheere, A. R. (2001). Storage and retrieval of addition facts: The role of number comparison. *The Quarterly Journal of Experimental Psychology Section A, 54*(4), 1005-1029. doi: 10.1080/713756007

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

Camos, V. (2008). Low working memory capacity impedes both efficiency and learning of number transcoding in children. *Journal of Experimental Child Psychology, 99*(1), 37-57. doi: 10.1016/j.jecp.2007.06.006

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., . . . Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science, 23*(6), 678-686. doi: 10.1177/0956797611434536

Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education, 15*(3), 179-202.

Castronovo, J., & Göbel, S. M. (2012). Impact of high mathematics education on the number sense. *PLoS ONE, 7*(4), e33832. doi: 10.1371/journal.pone.0033832

Chan, D. (2003). Data analysis and modeling longitudinal processes. *Group & Organization Management, 28*(3), 341-365. doi: 10.1177/1059601102250814

Cipolotti, L., & Butterworth, B. (1995). Toward a multiroute model of number processing: Impaired number transcoding with preserved calculation skills. *Journal of Educational Psychology: General, 124*(4), 375-390.

Dagenbach, D., & McCloskey, M. (1992). The organization of arithmetic facts in memory: Evidence from a brain-damaged patient. *Brain and Cognition, 20*(2), 345-366. doi: 10.1016/0278-2626(92)90026-I

Department for Education and Skills (DfES). (2006). Primary Framework for Literacy and Mathematics. London: DfES Publications.

De Smedt, B., & Gilmore, C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *Journal of Experimental Child Psychology, 108*(2), 278-292. doi: 10.1016/j.jecp.2010.09.003

De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology, 103*(4), 469-479. doi: 10.1016/j.jecp.2009.01.010

De Vos, T. (1992). *TTR: Tempotest rekenen*. Lisse, The Netherlands: Swets & Zeitlinger.

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

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*(3), 371-396.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition, 1*, 83-120.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex, 33*(2), 219-250. doi: 10.1016/s0010-9452(08)70002-9

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 626-641. doi: 10.1037/0096-1523.16.3.626

Dehaene, S., Izard, V., & Piazza, M. (2005). *Control over non-numerical parameters in numerosity experiments*. Retrieved from http://www.unicog.org/docs/DocumentationDotsGeneration.doc.

Delazer, M., & Benke, T. (1997). Arithmetic facts without meaning. *Cortex, 33*(4), 697-710. doi: 10.1016/s0010-9452(08)70727-5

Desoete, A., Ceulemans, A., De Weerdt, F., & 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*(1), 64-81. doi: 10.1348/2044-8279.002002

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

Dudal, P. (2000). *Leerlingvolgsysteem. Rekenen: Toetsen 1–2-3. Basisboek.* . Leuven, Belgium: Garant.

Dunn, L. M., Dunn, D. M., & Styles, B. (2010). *The British Picture Vocabulary Scale Third Edition (BPVS-III)*. London: GL Assessment.

Durand, M., Hulme, C., Larkin, R., & Snowling, M. (2005). The cognitive foundations of reading and arithmetic skills in 7- to 10-year-olds. *Journal of Experimental Child Psychology, 91*(2), 113-136. doi: 10.1016/j.jecp.2005.01.003

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307-314. doi: 10.1016/j.tics.2004.05.002

Fias, W., Brysbaert, M., Geypens, F., & d'Ydewalle, G. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition, 2*(1), 95-110. doi: 10.1080/135467996387552

Friso-van den Bos, I., van der Ven, S. H. G., Kroesbergen, E. H., & van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review, 10*, 29-44. doi: http://dx.doi.org/10.1016/j.edurev.2013.05.003

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*(3), 493-513. doi: 10.1037/0022-0663.97.3.493

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., . . . Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*(1), 29-43. doi: 10.1037/0022-0663.98.1.29

Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: contributions of inhibitory control. *Developmental Science, 16*(1), 136-148. doi: 10.1111/desc.12013

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*(1-2), 43-74. doi: 10.1016/0010-0277(92)90050-r

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: from reals to integers. *Trends in Cognitive Sciences, 4*(2), 59-65. doi: 10.1016/s1364-6613(99)01424-2

Geary, D. C., & Brown, S. C. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in gifted, normal, and mathematically disabled children. *Developmental Psychology, 27*(3), 398-406.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*(4), 1343-1359. doi: 10.1111/j.1467-8624.2007.01069.x

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition, 87*(3), B87-B95. doi: 10.1016/S0010-0277(02)00234-2

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., . . . Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS ONE, 8*(6), e67374. doi: 10.1371/journal.pone.0067374

Gilmore, C., Attridge, N., & Inglis, M. (2011). Measuring the approximate number system. *The Quarterly Journal of Experimental Psychology, 64*(11), 2099-2109. doi: 10.1080/17470218.2011.574710

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition, 115*(3), 394-406. doi: 10.1016/j.cognition.2010.02.002

Ginsburg, H. P., & Baroody, A. J. (1990). *Test of Early Mathematics Ability - Second Edition (TEMA-2)*. Austin, Texas: Pro-Ed.

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of Early Mathematics Ability - Third Edition (TEMA-3)*. Austin, Texas: Pro Ed.

Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of Experimental Child Psychology, 76*(2), 104-122. doi: 10.1006/jecp.2000.2564

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development, 58*(1), 80-92.

Grégoire, J., Noël, M.-P., & Van Nieuwenhoven, C. (2004). *TEDI-MATH*. Antwerpen: Harcourt.

Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review, 79*(4), 329-343. doi: 10.1037/h0032950

Haffner, J., Baro, K., Parzer, P., & Resch, F. (2005). *Heidelberger Rechentest (HRT)*. Göttingen, Germany: Hogrefe.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457-1465. doi: 10.1037/a0012682

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*(7213), 665-668. doi: 10.1038/nature07246

Hatcher, P. J., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development, 65*(1), 41-57. doi: 10.1111/j.1467-8624.1994.tb00733.x

Hecht, S. A., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computation skills: A longitudinal study from second to fifth grades. *Journal of Experimental Child Psychology, 79*(2), 192-227. doi: 10.1006/jecp.2000.2586

Hitch, G. J. (1978). The role of short-term working memory in mental arithmetic. *Cognitive Psychology, 10*(3), 302-323. doi: http://dx.doi.org/10.1016/0010-0285(78)90002-6

Holloway, I. D., & Ansari, D. (2008). Domain-specific and domain-general changes in children's development of number comparison. *Developmental Science, 11*(5), 644-649. doi: 10.1111/j.1467-7687.2008.00712.x

Holloway, I. D., & 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*(1), 17-29. doi: 10.1016/j.jecp.2008.04.001

Howell, D., C. (2010). *Statistical methods for Psychology* (7th ed.). Belmont, USA: Wadsworth Cengage Learning.

Hu, L., & 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), 1-55. doi: 10.1080/10705519909540118

Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F. J., & 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*(6), 572-577. doi: 10.1177/0956797611435921

Hulme, C., & Snowling, M. J. (2009). *Developmental disorders of language, learning and cognition*. Oxford, UK: Wiley-Blackwell.

Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin & Review, 18*(6), 1222-1229. doi: 10.3758/s13423-011-0154-1

Jou, J., & Aldridge, J. W. (1999). Memory representation of alphabetic position and interval information. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*(3), 680-701. doi: 10.1037/0278-7393.25.3.680

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin, 109*(3), 490-501.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *The American Journal of Psychology, 62*(4), 498-525.

Kiselev, S., Espy, K. A., & Sheffield, T. (2009). Age-related differences in reaction time task performance in young children. *Journal of Experimental Child Psychology, 102*(2), 150-166. doi: 10.1016/j.jecp.2008.02.002

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). London: The Guilford Press.

Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. M. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learning and Instruction, 25*, 95-103. doi: 10.1016/j.learninstruc.2012.12.001

Koponen, T., Salmi, P., Eklund, K., & Aro, T., Aro. (2013). Counting and RAN: Predictors of arithmetic calculation and reading fluency. *Journal of Educational Psychology, 105*(1), 162-175. doi: 10.1037/a0029285

Krajewski, K., & Schneider, W. (2009). Exploring the impact of phonological awareness, visual–spatial working memory, and preschool quantity–number competencies on mathematics achievement in elementary school: Findings from a 3-year longitudinal study. *Journal of Experimental Child Psychology, 103*(4), 516-531. doi: 10.1016/j.jecp.2009.03.009

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: a study of 8–9-year-old students. *Cognition, 93*(2), 99-125. doi: 10.1016/j.cognition.2003.11.004

Landerl, K., & Kölle, C. (2009). Typical and atypical development of basic numerical skills in elementary school. *Journal of Experimental Child Psychology, 103*(4), 546-565. doi: 10.1016/j.jecp.2008.12.006

Lembke, E., & Foegen, A. (2009). Identifying early numeracy indicators for kindergarten and first-grade students. *Learning Disabilities Research & Practice, 24*(1), 12-20. doi: 10.1111/j.1540-5826.2008.01273.x

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science, 14*(6), 1292-1300. doi: 10.1111/j.1467-7687.2011.01080.x

Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Is approximate number precision a stable predictor of math ability? *Learning and Individual Differences, 25*, 126-133. doi: 10.1016/j.lindif.2013.02.001

Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica, 141*(3), 373-379. doi: 10.1016/j.actpsy.2012.09.009

Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense: Large-number discrimination in human infants. *Psychological Science, 14*(5), 396-401. doi: 10.1111/1467-9280.01453

Lonnemann, J., Linkersdörfer, J., Hasselhorn, M., & Lindberg, S. (2011). Symbolic and non-symbolic distance effects in children and their connection with arithmetic skills. *Journal of Neurolinguistics, 24*(5), 583-591. doi: 10.1016/j.jneuroling.2011.02.004

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*(1), 149-157.

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition, 121*(2), 256-261. doi: 10.1016/j.cognition.2011.07.009

Maloney, E. A., Risko, E. F., Preston, F., Ansari, D., & Fugelsang, J. (2010). Challenging the reliability and validity of cognitive measures: The case of the numerical distance effect. *Acta Psychologica, 134*(2), 154-161. doi: 10.1016/j.actpsy.2010.01.006

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General, 111*(1), 1-22. doi: 10.1037/0096-3445.111.1.1

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011a). Impaired acuity of the approximate number system underlies mathematical learning disability. *Child Development, 82*(4), 1224-1237. doi: 10.1111/j.1467-8624.2011.01608.x

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011b). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE, 6*(9), e23749. doi: 10.1371/journal.pone.0023749

Mazzocco, M. M. M., & Thompson, R. E. (2005). Kindergarten predictors of math learning disability. *Learning Disabilities Research & Practice, 20*(3), 142-155. doi: 10.1111/j.1540-5826.2005.00129.x

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition, 44*(1–2), 107-157. doi: 10.1016/0010-0277(92)90052-J

McCloskey, M., Aliminosa, D., & Sokol, S. M. (1991). Facts, rules and procedures in normal calculation: Evidence from multiple single-patient studies of impaired arithmetic fact retrieval. *Brain and Cognition, 17*(2), 154-203. doi: 10.1016/0278-2626(91)90074-I

Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., & Nuerk, 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*(5), 1837-1851. doi: 10.1016/j.ridd.2011.03.012

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature, 215*, 1519-1520. doi: 10.1038/2151519a0

Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology, 103*(4), 490-502. doi: 10.1016/j.jecp.2009.02.003

Mussolin, C., Mejias, S., & Noël, M.-P. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition, 115*(1), 10-25. doi: 10.1016/j.cognition.2009.10.006

Mussolin, C., Nys, J., Leybaert, J., & Content, A. (2012). Relationships between approximate number system acuity and early symbolic number abilities. *Trends in Neuroscience and Education, 1*(1), 21-31. doi: 10.1016/j.tine.2012.09.003

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40*(5), 665-681. doi: 10.1037/0012-1649.40.5.665

Muthén, L. K., & Muthén, B. O. (1998-2011). *Mplus User's Guide. Sixth Edition.* Los Angeles, CA: Muthén & Muthén.

Neurobehavioral Systems. (Presentation) [Computer software]. Retrieved from http://www.neurobs.com

Nys, J., Ventura, P., Fernandes, T., Querido, L., Leybaert, J., & Content, A. (2013). Does math education modify the approximate number system? A comparison of schooled and unschooled adults. *Trends in Neuroscience and Education, 2*(1), 13-22. doi: http://dx.doi.org/10.1016/j.tine.2013.01.001

Ofsted. (n.d.). Inspection reports, from http://www.ofsted.gov.uk/inspection-reports/find-inspection-report

Parkman, J. M. (1971). Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology, 91*(2), 191-205. doi: 10.1037/h0031854

Passer, M. W., & Smith, R. E. (2007). *Psychology: The science of mind and behavior* (3rd ed.). New York: McGraw-Hill.

Passolunghi, M. C., Vercelloni, B., & Schadee, H. (2007). The precursors of mathematics learning: Working memory, phonological ability and numerical competence. *Cognitive Development, 22*(2), 165-184. doi: 10.1016/j.cogdev.2006.09.001

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., . . . Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition, 116*(1), 33-41. doi: 10.1016/j.cognition.2010.03.012

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron, 44*(3), 547-555. doi: 10.1016/j.neuron.2004.10.014

Piazza, M., Pica, P., Izard, V., Spelke, E. S., & Dehaene, S. (2013). Education enhances the acuity of the nonverbal approximate number system. *Psychological Science, 24*(6), 1037-1043. doi: 10.1177/0956797612464057

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science, 306*(5695), 499-503. doi: 10.1126/science.1102085

Power, R. J. D., & Dal Martello, M. F. (1990). The dictation of Italian numerals. *Language and Cognitive Processes, 5*(3), 237-254.

Price, G. R., Palmer, D., Battista, C., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychologica, 140*(1), 50-57. doi: 10.1016/j.actpsy.2012.02.008

Psychology Software Tools (E-Prime Version 1.1) [Computer software]. Retrieved from http://www.pstnet.com

Purpura, D. J., Baroody, A. J., & Lonigan, C. J. (2013). The transition from informal to formal mathematical knowledge: Mediation by numeral knowledge. *Journal of Educational Psychology, 105*(2), 453-464. doi: 10.1037/a0031753

Raven, J. C. (1998). *Raven's Standard Progressive Matrices Plus*. San Antonio, Texas: Harcourt Assessment Inc.

Reeve, R., Reynolds, F., Humberstone, J., & Butterworth, B. (2012). Stability and change in markers of core numerical competencies. *Journal of Experimental Psychology: General, 141*(4), 649-666. doi: 10.1037/a0027520

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science, 19*(6), 607-614. doi: 10.1111/j.1467-9280.2008.02130.x

Rousselle, L., Palmers, E., & Noël, M.-P. (2004). Magnitude comparison in preschoolers: what counts? Influence of perceptual variables. *Journal of Experimental Child Psychology, 87*(1), 57-84. doi: 10.1016/j.jecp.2003.10.005

Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology, 30*(2), 344-357. doi: 10.1111/j.2044-835X.2011.02048.x

Sasanguie, D., Defever, E., Van den Bussche, E., & Reynvoet, B. (2011). The reliability of and the relation between non-symbolic numerical distance effects in comparison, same-different judgments and priming. *Acta Psychologica, 136*(1), 73-80. doi: 10.1016/j.actpsy.2010.10.004

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number–space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology, 114*(3), 418-431. doi: 10.1016/j.jecp.2012.10.012

Sasanguie, D., Van den Bussche, E., & Reynvoet, B. (2012). Predictors for mathematics achievement? Evidence from a longitudinal study. *Mind, Brain, and Education, 6*, 119-128. doi: 10.1111/j.1751-228X.2012.01147.x

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools Inc.

Schneider, W., Roth, E., & Ennemoser, M. (2000). Training phonological skills and letter knowledge in children at risk for dyslexia: A comparison of three kindergarten intervention programs. *Journal of Educational Psychology, 92*(2), 284-295. doi: 10.1037/0022-0663.92.2.284

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development, 48*, 630-633.

Seron, X., & Fayol, M. (1994). Number transcoding in children: A functional analysis. *British Journal of Developmental Psychology, 12*(3), 281-300. doi: 10.1111/j.2044-835X.1994.tb00635.x

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116*(3), 250-264. doi: 10.1037/0096-3445.116.3.250

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General, 117*(3), 258-275.

Siegler, R. S. (1989). Hazards of mental chronometry: An example from children's subtraction. *Journal of Educational Psychology, 81*(4), 497-506.

Siegler, R. S., & Shrager, J. (1984). Strategy choice in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229-293). Hillsdale, NJ: Erlbaum.

Soltész, F., Szűcs, D., & Szűcs, L. (2010). Relationships between magnitude representation, counting and memory in 4- to 7-year-old children: A developmental study. *Behavioral and Brain Functions, 6*(1), 13. doi: 10.1186/1744-9081-6-13

van Harskamp, N. J., & Cipolotti, L. (2001). Selective impairments for addition, subtraction and multiplication. Implications for the organisation of arithmetical facts. *Cortex, 37*(3), 363-388. doi: 10.1016/s0010-9452(08)70579-3

van Opstal, F., Gevers, W., De Moor, W., & Verguts, T. (2008). Dissecting the symbolic distance effect: Comparison and priming effects in numerical and nonnumerical orders. *Psychonomic Bulletin & Review, 15*(2), 419-425. doi: 10.3758/pbr.15.2.419

van Opstal, F., & Verguts, T. (2011). The origins of the numerical distance effect: The same–different task. *Journal of Cognitive Psychology, 23*(1), 112-120. doi: 10.1080/20445911.2011.466796

Vanbinst, K., Ghesquière, P., & De Smedt, B. (2012). Numerical magnitude representations and individual differences in children's arithmetic strategy use. *Mind, Brain, and Education, 6*(3), 129-136. doi: 10.1111/j.1751-228X.2012.01148.x

von Aster, M. G., & Shalev, R. S. (2007). Number development and developmental dyscalculia. *Developmental Medicine & Child Neurology, 49*(11), 868-873. doi: 10.1111/j.1469-8749.2007.00868.x

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, Texas: Harcourt Assessment Inc.

Wechsler, D. (2005). *Wechsler Individual Achievement Test - Second UK Edition (WIAT-II UK)*. London: Harcourt Assessment Inc.

Werner, C., & Schermellah-Engel, K. (May 2009). Structural Equation Modeling: Advantages, Challenges, and Problems. Retrieved 29 August, 2012, from http://www.psychologie.uzh.ch/fachrichtungen/methoden/team/christinawerner/sem/sem_pro_con_en.pdf

Westwood, P., Harris-Hughes, M., Lucas, G., Nolan, J., & Scrymgeour, K. (1974). One minute addition test - one minute subtraction test. *Remedial Education, 9*(2), 70-72.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science, 10*(2), 130-137. doi: 10.1111/1467-9280.00120

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson Psycho-Educational Battery-Revised*. Allen, Texas: DLM Teaching Resources.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III Tests of Achievement*. Itasca, IL: Riverside.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358*, 749-750. doi: 10.1038/358749a0

Wynn, K. (1998). Numerical competence in infants. In C. Donlan (Ed.), *The Development of Mathematical Skills* (pp. 3-25). Hove: Psychology Press.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*(1), B1-B11. doi: 10.1016/S0010-0277(99)00066-9

Zorzi, M., & Butterworth, B. (1999). A computational model of number comparison. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society* (pp. 772-777). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Zorzi, M., Stoianov, I., & Umilta, C. (2005). Computational modelling of numerical cognition. In J. I. D. Campbell (Ed.), *Handbook of Mathematical Cognition* (pp. 67-84). Hove: Psychology Press.