NUMERICAL BUILDING BLOCKS:

EXPLORING DOMAIN-SPECIFIC COGNITIVE PREDICTORS OF MATHEMATICS

by

Carla Sowinski

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Psychology

Carleton University
Ottawa, Ontario

©2016
Carla Sowinski

Abstract

This dissertation includes three studies examining individual differences in domain-specific quantitative skills as predictors of adults' mathematical performance. Quantitative skills included subitizing, counting, approximate number system (ANS), and symbolic skills. Subitizing is the ability to quickly and exactly enumerate small sets without counting (1 to 3 or 4), whereas the ANS facilitates discrimination between large quantities. Given the evidence for their presence among human infants and other animals, the subitizing and approximate number systems are considered core quantitative systems—leading to theories that one or both systems scaffold the acquisition of symbolic quantity representations. Counting is the process of enumerating sets beyond the subitizable range to determine exact quantity; learning to count is the first step in acquiring the symbolic system. The present research was framed by three theoretical accounts, each of which emphasize the subitizing, counting, *or* approximate number system as the key contributor to mathematical success. Compared to the ANS literature, very little research has examined subitizing and counting skills in relation to mathematics performance with adult samples. To address this issue, the current research included subitizing, counting, and ANS—as well as symbolic skills. These domain-specific quantitative skills were examined in relation to each other and as relative contributors to mathematics outcomes via path analyses (Studies 1 and 2) and structural equation modeling (Study 3). ANS skill did not uniquely predict mathematical outcomes requiring exact calculation, but did predict symbolic and nonsymbolic number line performance. Counting predicted symbolic quantitative skills, but not mathematical outcomes. Subitizing emerged as a predictor of *arithmetic fluency* across all three studies, but did not predict other mathematical outcomes. As hypothesized, symbolic quantitative skill tended to be the strongest predictor of all mathematical outcomes, except for nonsymbolic number line. Experiential factors also predicted mathematical outcomes across all three studies. These findings suggest that the subitizing system scaffolds the development of counting and symbolic quantitative skills, and continues to predict arithmetic fluency in adulthood. It is recommended that future research explore the role of subitizing in the development of symbolic quantitative skills, to gain understanding of this developmental trajectory.

Acknowledgements

Many thanks to my graduate advisor, Jo-Anne LeFevre. When I applied to graduate school, I had no idea how lucky I was be to be accepted by a supervisor like you. I am so glad that I ended up in the field of mathematical cognition, which has proved to be both challenging and fruitful. It’s amazing to think about how much I have learned and experienced since I entered the LeFevre Mathlab. Jo-Anne: You are such a wonderful mentor. You are so hard-working, but in my opinion the best thing about you is that you understand how to foster the talents of your graduate students. You encourage us to excel and be the best that we can be, but you also display so much empathy and caring when we need it. Best academic mom ever! I may be leaving the SS Mathlab, but you will always be my Captain Picard!

Thank you John Zelenski: My academic journey at Carleton University started and ended with you. Thank you to Deepthi Kamawar who taught me about cognitive development and contributed to both my MA and PhD theses. Thank you to Janet Mantler for supporting and encouraging me over the years. Thank you to Andrea Howard for the invaluable statistical advice with regards to structural equation modeling.

To Erica: You are one of my oldest friends and my first academic friend, as you were the first person that I knew that went to graduate school. Thank you for years of friendship, and for the chocolate-covered coffee beans and encouraging jelly fish to help me through the final stages of my thesis. The jelly fish cheered me on from the wall beside the desk where I was writing. To Deanna Whelan: You are my longest-standing “psychology” friend. Thank you for encouraging me to go to grad school; you believed in me when I wasn’t so sure. And thanks for all of the coffee dates!

Thank you to my true friends near and far. Thank you to my dancing friends for helping me unwind and for bringing joy to my life: Francine Beaulne, Christine Kemp, Chelsea, Cheryl Harasymchuk, Emilie, Francine L., Charlie, Jaime, & Yvon. To the Sarnia crew, Sunil, Kurt, Chris, Ross, Laurissa, & Jason: I love you guys; I'm glad we've been friends for so long and I hope that doesn't change anytime soon. Thank you to my ever-supportive Ottawa friends for being my family away from home: Anastasia, Ben, Cory, Nicole Robert, Heather, Michael & Rebecca Windover, Viktor, Frank, Natalie, and the remarkable Re-Cycles Bike Shop people.

A special thank you to my best friend, Nicole Aitken. Nicole and I TAd together in our first semester of grad school. It's been great to be on this journey with you, Nicole! You are my favourite friend to travel with (and do all other things, too). I love it that you are up for just about anything. I'm so grateful to have a friend like you, as I never had one that was so awesome before! You're always so supportive—and this was particularly true during the last phase of this thesis. I hope we have many more adventures together in the future! Love you!

To mom and dad, Pat and Ron: I love you guys so much. Thank you for always being so supportive—even if, at times, you were a little unsure of what all this research stuff is anyway. They say that we reach great heights by standing on the shoulders of giants. You gave me the strong foundation that I needed to accomplish this PhD. Even though I don't see you that much, I think of you often. Your kindness and generosity are with me always. I know you are proud, and it means a lot to me.

To my sisters, Sarah, Theresa, Jessica and Jenna: Thank you for your support. Each of you sent encouraging messages and told me you were proud of me. I love that we all come from the same place, but we are all so unique. I love you, sisters! Thank you to my inlaws, Elaine and Tony, for your support, too. I always enjoy getting to spend time with you—perhaps over tea. Thank you to my grandparents, Bill and Elly Gilroy. You came to every birthday party, recital, school assembly, and picked me up from school when I was sick. It wasn't until I was older that I really came to appreciate how lucky I am to have grandparents like you.

Derek: Thank you for being supportive throughout this degree, and throughout our marriage. You always believe in me, even when I am unsure. Your unconditional love is something that I cherish. Thanks for thinking I'm the best. Thanks for the little things that mean a lot (e.g., coffees, Caesar salads, candy, rides, computer help, Bill Nye swag—the list goes on). You are always thinking of me, and you always help me when I ask. I know it was hard for you when I was working so much on this big project, so thank you also for your patience and understanding. Finally, thanks to my cats, Rascal, Hawkeye, and Trapper for providing ongoing entertainment and for hanging out with me late at night when I was writing.

I'm so grateful to have had this incredible opportunity.

Table of Contents

List of Tables

List of Figures

List of Appendices

CHAPTER 1: INTRODUCTION

Mathematics knowledge is important to academic achievement and career success (e.g., Bynner & Parsons, 1997). In 2007, Duncan et al. published a groundbreaking meta-analysis demonstrating that school-entry mathematics readiness skills (e.g., ordinality; number knowledge) were stronger predictors of later academic achievement than literacy and attention skills. In contrast, school-entry social emotional factors were not predictive of academic success. These results were surprising because up until this point, early literacy and socioemotional skills were believed to be the most crucial to academic achievement. These findings emphasized the need to understand mathematical development in order to promote not only subsequent mathematic skill, but also overall achievement—thus highlighting the importance of research in this field.

Despite a surge of interest in mathematical cognition (De Smedt, Noel, Gilmore, & Ansari, 2013), there remains a lack of understanding about the cognitive precursors and experiential factors that support mathematical learning—especially in comparison to what is known about factors that predict literacy acquisition. For example, evidence suggests parents' activities, specifically shared storybook reading, contribute to children's acquisition of vocabulary and reading comprehension skills (Sénéchal, 2006; Sénéchal, & LeFevre, 2002; Sénéchal, Pagan, Lever & Ouellette, 2008; Sénéchal & Young, 2008). Decades of research has shown that children with good phonological awareness learn to read more easily because phonological skills help them learn the mappings between letters and sounds (i.e., orthography and phonology; e.g., Bryant, MacLean, Bradley & Crossland, 1990; Lonigan, Burgess, Anthony & Barker, 1998; Melby-Lervag, Lyster, & Hulme, 2012; Treutlein, Zoller, Roos, & Scholer, 2009;

Wagner & Torgeson, 1987). In sum, there is a general consensus that vocabulary, reading comprehension, and phonological awareness are building blocks of literacy, and practices to promote these skills are well-known. In contrast, the building blocks of mathematics and ways to promote mathematical development remain poorly defined (cf. Skwarchuk, Sowinski, & LeFevre, 2014; Sowinski et al., 2015), despite considerable research in this area over the last 15 years. Thus, the present research explored domain-specific quantitative skills as predictors of mathematical outcomes among adults, with the goal of gaining some understanding of the building blocks of mathematics.

While domain-general abilities, like working memory and executive function, are predictive, they are not sufficient to fully explain individual differences in mathematical achievement (Butterworth, 2005, 2010; Landerl, Bevan & Butterworth, 2004). This has led researchers to posit that domain-specific quantitative abilities contribute to such success. To explore which (if any) domain-specific predictors might account for individual differences in mathematical skill, quantitative skills were selected based on the existing theoretical accounts and research evidence. These quantitative skills were examined as unique predictors of mathematical outcomes with adult samples across three studies. This thesis is organized as follows: The present chapter includes a review of the relevant literature to justify and operationalize the selected domain-specific quantitative skills and domain-general cognitive skills (Study 3 only). Chapter 2 contains Studies 1 and 2; Chapter 3 contains Study 3. Within each study, I describe the research goals and hypotheses, methodology, results, and discuss the findings. Chapter 4 contains the general discussion, including recommendations for future research and conclusions.

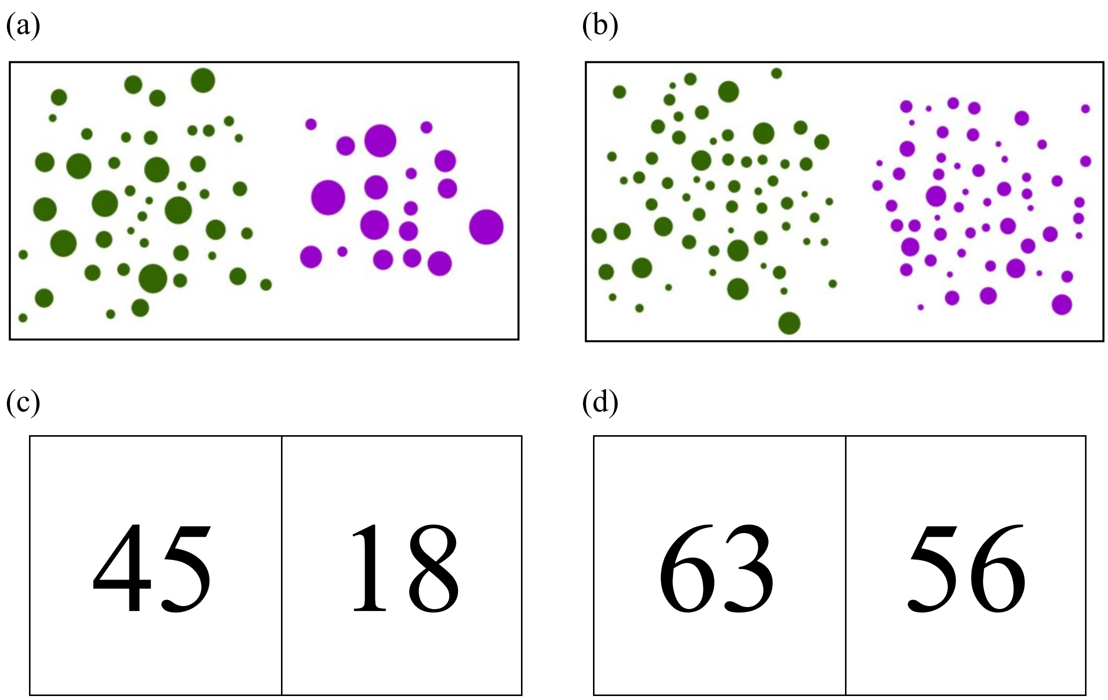
Literature Review

In this chapter, I first describe two core quantitative systems—the subitizing and the approximate number systems, followed by a description of the counting system. Second, I define three ranges of quantities—small (1 to 3), medium (5 to 9), and large (10 to 100). Third, I describe the common methods of indexing the subitizing, counting, and approximate number systems; I argue for the inclusion of identification, comparison, and ordering tasks in the present research. Fourth, I review three theoretical accounts that make different claims about a) how symbolic quantitative skills become attached to nonsymbolic quantity processes, and b) the importance of the subitizing, counting, and ANS systems for mathematical success. The evidence to support or refute these accounts is discussed as well, thus providing the theoretical background for this thesis. Fifth, I review the evidence on symbolic and nonsymbolic quantitative skills in relation to math performance as rationale for including analogous tasks in both formats. Finally, I provide rationale for the inclusion of domain-general and experiential factors.

Core Quantitative Systems

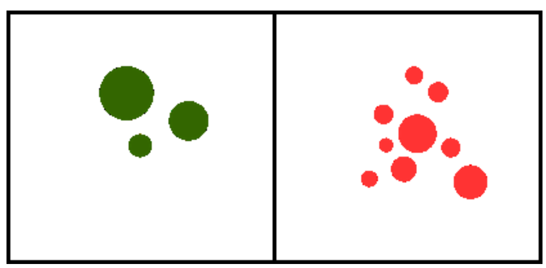
Most researchers agree that there are two core quantitative systems: the approximate number system (ANS) and the subitizing system (also known as the parallel individuation or object tracking system; Butterworth, 2010; Cutini, Scatturin, Basso Moro, & Zorzi, 2014; Feigenson, Dehaene & Spelke, 2004; Hyde, 2011; Izard, Pica, Spelke & Dehaene, 2008; Mou & vanMarle, 2014; Noël & Rousselle, 2011; Piazza et al., 2010). The ANS enables us to discriminate between large quantities (Feigenson et al., 2004; Halberda, Mazzocco & Feigenson, 2008; Mazzocco, Feigenson & Halberda, 2011a). This ability is ratio dependent, which means discrimination is easy when the difference between quantities is large, as in Figure 1.1(a), (e.g., 45 versus 18 dots; ratio of 5:2), but becomes more difficult as the ratio approaches 1:1, as in Figure 1.1(b) (e.g., 63 versus 56 dots; ratio of 9:8). The ratio effect suggests this quantitative system provides an approximate (i.e., noisy) representation of quantity. The quantity representations are not exact, and they overlap with nearby quantities (see Feigenson et al., 2004 for a more in-depth discussion). This ratio effect follows Weber's law which states that the ability to detect a difference between two stimuli is proportional to the size (or intensity) of the stimuli. This law applies to many features besides quantity, such as loudness and brightness (Leibovich & Ansari, 2016)

Figure 1.1. Example ANS discrimination stimuli (panels a and b) with analogous symbolic discrimination shown below (panels c and d). It is easier to discriminate between the two quantities in the example on the left (ratio of 5:2) in comparison to the example on the right (ratio of 9:8).



In contrast, subitizing is an *exact* (rather than approximate) system that has been implicated in humans' ability to keep track of small sets of objects (i.e., up to 3 or 4; Clements, 1999; Feigenson et al., 2004; Mandler & Shebo,1982; Trick, 2008; Trick & Pylyshyn, 1994). Most adults can quickly and easily enumerate 1 to 3 or 4 objects without having to count, however, each object needs to be counted to determine exact quantities of 5 or greater. The subitizing effect is thought to originate in a domain-general attentional system for tracking small sets of objects (Trick & Pylyshyn, 1994). In contrast to subitizing, counting is a much more effortful process and it takes more time (Mandler & Shebo, 1982; Trick, 2008; Trick & Pylyshyn, 1994). See Figure 1.2 for a comparison of quantities that can be subitized (3) versus counted (9).

Figure 1.2: Example of nonsymbolic quantities. The quantity in the left panel can be subitized and the one on the right must be counted.



The ANS is presumed to be a core quantitative system based on its observed presence in human infants and other animals. Using a habituation paradigm, six-month-old infants can distinguish between two large quantities when the ratio between them is 1:2 (e.g., 8 versus 16; Xu & Spelke, 2000), and in some cases (e.g., audiovisual events) when the ratio is 2:3 (Jordan, Suanda, & Brannon, 2008). By 9 months of age, infants are able to discriminate 8 objects from 12 (ratio of 2:3; Lipton & Spelke, 2003). These results hold when researchers control for total surface area and when other stimulus modalities, such as sounds, are used, suggesting that infants are sensitive to quantity and are not just responding based on their visual or perceptual systems (Feigenson et al., 2004; Lipton & Spelke, 2003; Xu & Spelke, 2000). The acuity of the approximate number system improves across development (Gilmore, McCarthy, & Spelke, 2010; Halberda & Feigenson, 2008; Piazza et al., 2010) such that the average adult can discriminate between quantities up to ratios of 7:8 (Mundy & Gilmore, 2009) or 9:10 (Pica, Lemer, Izard & Dahaen, 2004; Halberda & Feigenson, 2008). There is also evidence for the presence of an approximate number system in other animals, such as bird, lizards, rats, and monkeys (Ansari, 2008; Brannon, 2005; Dehaene, 1997; Dehaene, Dehaene-Lambertz & Cohen, 1998; Feigenson et al., 2004). In sum, the ANS is a noisy, ratio-dependent number representation system present in human infants, children, adults, and many other animals. The accuracy of this system improves across development, but retains a ratio limit.

Recognition of small exact quantities, or *subitizing*, is also presumed to be a core quantitative system. Human infants can track small sets of individual objects (1 to 3; Feigenson, Carey & Spelke, 2002; Feigenson et al., 2004). Using a dishabituation paradigm, Starkey and Cooper (1980) found that 22-week-old infants were able to distinguish 2 from 3, but not 4 from 6. Feigenson, Carey and Hauser (2002) had 10- and 12-month-old infants watch as two containers holding crackers were placed in a room. Infants then crawled to select a container. They chose the larger quantity of crackers when comparisons were 1 versus 2, and 3 versus 1, but chose randomly for comparisons of 2 versus 4, 3 versus 4, and 3 versus 6. These findings suggest that babies are able to track individual items up to a maximum of 3 objects.

Feigenson and Carey (2003) explored babies' abilities to track objects in a manual search paradigm. In Experiment 1, when there were 3 or fewer items, babies (aged 14.5 months) searched longer for items when more items were expected. For example, when 2 items were hidden in a box, and only 1 item had been retrieved, babies spent more time trying to retrieve the extra item in comparison to trials in which no other items were expected (e.g., when 1 item was hidden and 1 item had been retrieved). Babies searched for an additional item when 3 items were inserted and 2 retrieved, but not when 4 items were inserted and 2 retrieved. These findings suggest that babies were able to track 3 but not 4 objects. In fact, they were not even sensitive to four as *more* than two, as searching for a third object would have resulted in success on this task (see also Experiment 1 of Feigenson & Carey, 2005). In Experiment 2, Feigenson and Carey (2003) varied the size of the objects retrieved to see if babies (M_{age} = 12.5 months) would end their search when one large object was retrieved in place of two smaller objects. They found that babies continued their search, indicating sensitivity to exact quantity beyond a sensitivity to overall volume or area of the objects.

There is also evidence that animals, such as monkeys, can subitize. Hauser, Carey and Hauser (2000) conducted an experiment with rhesus macaques. The macaques watched as different quantities of apple slices were placed in boxes (e.g., 1 vs. 2); the box approached first was taken as the monkey's choice. They were successful at selecting 2 versus 1, 3 versus 2, and 4 versus 3, but performed at chance on comparisons of 4 versus 5, 4 versus 6, 4 versus 8, and 3 versus 8. Thus, monkeys were able to track up to four objects (versus three in babies), but were unable to track more than this amount; monkeys performed at chance even on the larger comparisons with a large ratio difference (e.g., 3 versus 8). Thus, similar to the research with infants, monkeys seem to be able to track a small number of individual objects, but do not exhibit the ratio effect associated with the ANS with (relatively) small quantities (e.g., 4 versus 8).

In summary, the subitizing and approximate number systems are argued to be core quantitative systems that are evolutionarily primitive, given the evidence with non-human primates, and innate, given the evidence with animals and human infants. The ANS is responsible for discrimination between *large* numerosities and is characterized by the ratio effect, which suggests that its quantity representations are approximate in nature. In contrast the subitizing system is responsible for the tracking and quick apprehension of small exact quantities (Feigenson et al., 2004). Neither system, however, is useful for exact enumeration of quantities greater than four. Humans have surmounted this limitation by developing the counting system that combines language-based and quantitative knowledge to allow exact enumeration of any quantity.

The Counting System

Counting is the sequential labeling of objects with the goal of determining “How many?”. It represents the transition from the use of nonverbal, nonsymbolic quantitative representations to the use of verbal and symbolic number representations. To the average adult, counting may seem like a simple task, but this apparent simplicity is likely due to years of practice. Even among adults, counting is a demanding process—one that requires working memory resources (Hecht, 2002).

Before children fully understand how to count, they must acquire specific count principles (Gelman and Gallistel, 1978). Children begin counting around age 2 years, and counting skills improve slowly between ages 2 to 4 years (Butterworth 2005, 2010; Wynn, 1990). First an arbitrary string of sounds must be learned—the count words—which have no meaning on their own. Children learn that these words have a *stable order* (Gelman & Gallistel, 1978). *One-to-one correspondence* must also be learned, which means that each number symbol (i.e., counting word) is linked to only one object; no number word is repeated, and all objects must be counted. Children must learn that the last number in the count sequence represents the number of objects in the set—known as the *cardinality* principle. Once a child grasps the cardinality principle s/he has some understanding of the purpose of the count system (i.e., determining "how many?"). However, cardinality knowledge is gained incrementally. Children aged 2 to 3.5 years can typically recite the number sequence up to ten, but can only map a few of these number words to their cardinal values; these children are known as "subset-knowers". To assess their knowledge of the cardinality principle, children are asked to produce a set of a specific quantity, known as the Give-N task (e.g., "Can you give me 2 apples?"; Wynn, 1990). A child that is a subset-knower can accurately produce sets within his/her set knowledge, but beyond that gives variable answers. For example a *two-knower* can produce sets of one and two, but if asked for any set greater than two, would likely give a handful of objects (but—importantly—*not* one or two objects). Eventually children can produce sets as large as they can count verbally, and come to understand that the last word spoken when counting represents the number of objects in the set. At this point they are *cardinality principle (CP)-knowers* (Izard, Streri & Spelke, 2014; Wynn, 1990).

Children must also learn that anything can be counted, known as the *abstractness* principle, and that the order of the count does not matter, known as the *order-irrelevance* principle (e.g., objects can be counted in any order; Gelman & Gallistel, 1978). These last two principles take longer to develop. In fact, rejection of the order-irrelevance principle is linked to math achievement in young children, perhaps because counting objects in order helps children perform more accurately until they have built up their proficiency (see Kamawar et al., 2010; LeFevre et al., 2006). To summarize, learning how to count and execute counting procedures correctly are demanding and effortful tasks for young children (Feigenson et al., 2004; Gelman & Butterworth, 2005; Le Corre, Van de Walle, Brannon & Carey, 2006). In contrast to subitizing and discriminating between large quantities, via the ANS, counting has to be learned.

Three Distinct Systems in Small, Medium, and Large Ranges

It is important to discuss the evidence that the subitizing, counting, and approximate number systems are distinct systems; this evidence serves as rationale for the three quantity ranges defined in the current thesis. Subitizing operates only on small quantities whereas ANS processing has usually been associated with large quantities. Nevertheless, some researchers have argued that small and large nonsymbolic quantitative skills employ a single cognitive system—the ANS (Cordes, Gelman, Gallistel & Whalen, 2001; Dehaene & Changeux, 1993; Gallistel & Gelman, 1991).

Revkin, Piazza, Izard, Cohen and Dehaene (2008) tested the assumption that the ANS applies to small and large quantities with an adult sample. They found that accuracy on small nonsymbolic comparison was much better than large, even when ratios were the same (e.g., 1 vs. 8 in comparison to 10 vs. 80 dots), suggesting that small number processing is different and distinct from large number processing. The evidence with babies and animals also supports the view that subitizing and approximate number systems are separate. Babies did not discriminate 2 versus 4 or 3 versus 6 objects (Feigenson, et al., 2002), even though they discriminated at the same ratio of 1:2 when quantities were larger (e.g., 8 versus 16; Xu & Spelke, 2000). Similarly, in Hauser et al.'s (2000) research, monkeys failed to discriminate 4 from 8, but succeeded when quantities were 4 and fewer. More recently, research examining event-related potentials (ERPs) found different behavioural signatures for small versus large numbers (Hyde & Spelke, 2009). Cordes and Brannon (2009) also found that babies were unable to discriminate between quantities that crossed the subitizing range (e.g., 2 versus 6) until the ratio became 1:4 (e.g., 2 versus 8). The neuroimaging research is also inconsistent with a one-system view (see Hyde 2011 for a review). These findings support the claim that the subitizing and ANS systems are *distinct* pre-symbolic quantity systems.

These findings also suggest that some quantities are too large for the subitizing system, but not yet large enough to employ the ANS, which is why the present research includes a medium range (5 to 9) between the small (tapping subitizing) and large (tapping ANS) ranges. Nonsymbolic tasks in the medium range were presumed to tap the counting range. Note that Noël and Rousselle (2011) posit that the counting system is particularly important for the development of math skill (as described in the next section); by including separate counting measures this hypothesis can be tested. In sum, the creation of three distinct quantity ranges was based on the evidence for distinct quantitative systems. The inclusion of three ranges permits hypothesis testing about the relative importance of subitizing, counting, and ANS systems to mathematics skill.

Accordingly, in order to clarify the relations between number size and number system, the present research includes the quantities 1 to 100 in symbolic and nonsymbolic formats (i.e., dots and digits). The small range includes quantities 1 to 3; nonsymbolic tasks in this range are presumed to index the subitizing system. The quantity 4 was not included in the final scoring of tasks as there are differences in individual subitizing ranges (Trick & Pylyshyn, 1994); excluding 4 increases the likelihood that the majority of participants in the present research were subitizing on nonsymbolic tasks in the small range—that is quickly and exactly determining the quantity. The medium range included quantities 5 to 9; nonsymbolic tasks in this range were presumed to tap the counting system. The large range included quantities 10 to 100; nonsymbolic tasks in this range were presumed to index the ANS.

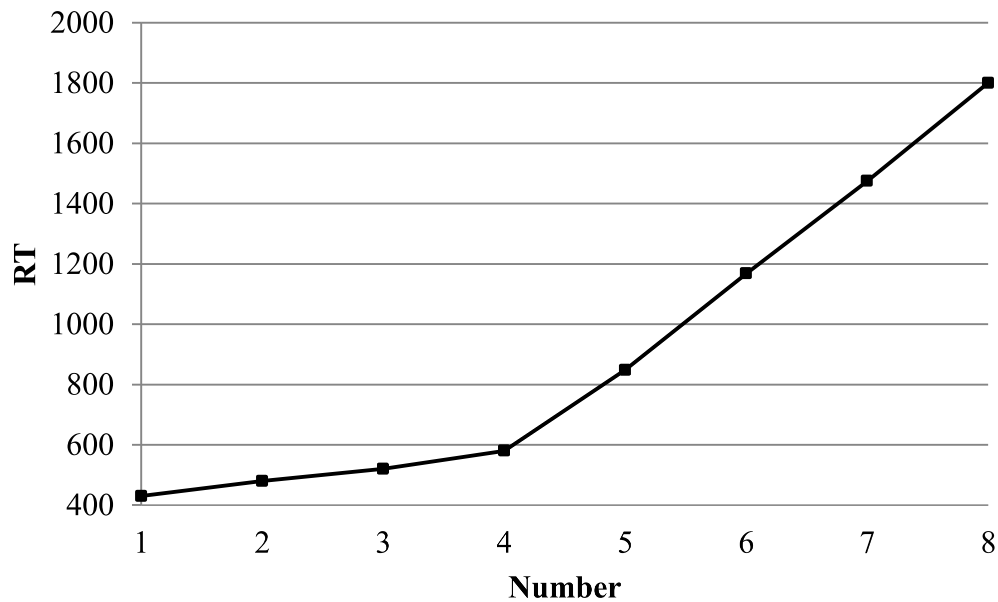
Measuring the Subitizing, Counting, and Approximate Number Systems

Before describing the relevant theory and research, it is important to describe the main ways that subitizing, counting, and ANS systems have been assessed among adults.

Subitizing and counting. It is common for a single identification or enumeration task to index both subitizing and counting skill (see Trick & Pylyshyn, 1994). For example participants might be shown 1 to 5 (e.g. Gray & Reeve, 2014) up to 9 dots (e.g., Lyons, Price, Vaessen, Blomert, & Ansari, 2014; Piazza, Mechelli, Butterworth & Price, 2002) and asked to state "How many?". Objects remain visible until a response is given. Subitizing versus counting ranges are evident when RTs are plotted by quantity; see Figure 1.3. Subitizing is characterized by a fairly flat slope within the 1 to 3 or 4 range (40 to 100ms per item), suggesting quick and exact apprehension of these small quantities. A much steeper slope is typically observed for quantities 5 (or 4) and greater

(250 to 350ms per item; Trick & Pylyshyn, 1994). The point of discontinuity in the reaction time slope is taken as evidence of a shift from subitizing to counting (Schleifer & Landerl, 2010). Thus, subitizing performance can be indexed with the mean or median RTs in the subitizing range, the slope of the subitizing range, and/or the span of the subitizing range (some people show a fairly flat slope over a larger range than others). Counting is typically indexed by mean RT or slope within the counting range.

Figure 1.3. A typical RT graph for an enumeration task with quantities 1 to 9.



Approximate number system. In a recent paper, Dietrich, Huber and Nuerk (2015b) reviewed the methodological aspects of various ANS tasks, and provided task recommendations based on the evidence. Among adults, ANS acuity is typically assessed via computerized dot comparison tasks (Dietrich et al., 2015b; Schneider et al., 2016); squares, triangles, or other objects may be used, but dots are most common. Participants view two sets of dots and select the one that is greater in numerosity. Dot arrays may be presented sequentially, or simultaneously—either side-by-side (paired) or intermixed (e.g., blue and yellow dots intermixed; Dietrich et al., 2015b). Same-different tasks (i.e., participant is shown two sets and decides if they are the same or different), and approximate arithmetic tasks have also been used to measure the ANS. Dietrich et al. (2015b) argue that paired presentation is the best format because it does not require additional domain-general cognitive abilities, like visual resolution (perhaps involved in intermixed comparison) or working memory (probably involved in sequential comparison or approximate arithmetic; Gilmore, Attridge, De Smedt, & Inglis., 2014; Price, Palmer, Battista, & Ansari, 2012), and thus results in the "purest" measure of ANS (i.e., discriminant validity is maximized). Paired dot comparison is the most frequently employed ANS task, and the different task-types are not *highly* correlated (see Gilmore, Attridge, & Inglis., 2011, 2014; Price et al., 2012; Smets, Sasanguie, Szűcs, & Reynvoet, 2015), thus another reason to select paired dot comparison is to maximize consistency with other research findings (i.e., replicability). The paired dot comparison task has also demonstrated adequate reliability (Gilmore et al., 2011, 2014; Price et al., 2012).

The range of the quantities in the dot arrays is also an important consideration. Because the ANS and subitizing systems have been identified as two distinct numerical systems, with different behavioural signatures, a task intended to measure the ANS should not include quantities in the subitizable range (i.e., 1 to 4)—or quantities just above it as these may be counted (Dietrich et al., 2015b). With that in mind, the dot arrays in ANS tasks should be presented briefly—the assumption being that brief presentation leads to activation of the approximate number system and prevents counting. Paired dot tasks should include pairs from a variety of ratios so as to avoid floor or ceiling effects and ensure variability in ANS performance (Dietrich et al., 2015b).

In recent years, researchers have identified visual characteristics of stimuli that are associated—and thus confounded—with numerosity (see Gebuis & Reynvoet, 2011 for recommendations on creating nonsymbolic magnitude comparison stimuli). The relevant confounds are briefly described here. If dots in the two arrays are the same size, the array that is greater in numerosity will also occupy greater overall area, thus participants could succeed based on area rather than numerosity. The overall area of the two arrays can be made equal, but if the dots *within* each array are the same size, the greater numerosity is always comprised of smaller dots, creating another consistent cue that might aid performance. To circumvent this issue, researchers have equated area within each pairing and varied the dot sizes within arrays (Ansari & Dhital, 2006; Izard, Sann, Spelke & Streri, 2009; Libertus, Woldorff, & Brannon, 2007; Piazza, Pinel, Le Bihan, & Dehaene, 2007; Santens, Roggerman, Fias, & Verguts, 2010). Gebuis and Reynvoet (2011, 2012) argued that dot density and perimeter, that is, the band of area in which the dots appear (also known as convex hull) should also be taken into consideration. For example, the smaller set of dots should not be consistently grouped in the middle of the stimulus area, but instead, at times, should appear within the same perimeter as the larger set; similarly, larger quantities should not consistently appear more densely packed than smaller quantities. At present it is common practice for a) area to be correlated with numerosity on half of trials, b) area to be *negatively* correlated with numerosity and perimeter-equated for the other half of trials (i.e., the smaller array occupies more total surface area, but the spread of two arrays is the same), and c) dots to appear in a variety of sizes within each array. In sum, it is recommended that ANS measures vary multiple visual cues within the dot comparison task to encourage performance based on numerosity rather than visual cues.

A variety of performance measures can be derived from nonsymbolic magnitude comparison tasks, including numerical distance effect (NDE), numerical ratio effect (NRE) accuracy, response time (with or without a correction for accuracy), or by calculating a Weber fraction; see Inglis and Gilmore's article on indexing the ANS (2014). Many researchers have used the Weber fraction (*w*) to index individual differences in approximate number system acuity (Halberda et al., 2008; Lyons & Beilock, 2009, 2011; Mazzocco et al., 2011a; Pica et al., 2004; Revkin et al., 2008). The Weber fraction represents the "minimal differences in intensity between stimuli that can still be discriminated" (p. 13, Leibovich & Ansari, 2016; Mazzocco et al., 2011a; Mazzocco, Feigenson & Halberda, 2011b); higher Weber fractions reflect lower approximate number system acuity (Leibovich & Ansari, 2016; Mazzocco et al., 2011a). However, as demonstrated in Inglis and Gilmore (2014) the Weber fraction has a tendency to have a skewed distribution and has less-than-acceptable test-retest reliability. Calculating numerical distance effect as an index of the ANS was also not recommended as it demonstrated extremely low test-retest reliability (Inglis & Gilmore, 2014; Lindskog, Winman, Juslin, & Poom, 2013; Sasanguie, Defever, Van den Bussche, & Reynvoet, 2011). The numerical ratio effect (NRE) is not recommended either, as it may not correlate with other ANS performance measures (Inglis & Gilmore, 2014; Lindskog et al., 2013). Response times and error rates have also been used as performance measures for quantity comparison tasks (e.g., Holloway & Ansari, 2009; Libertus, Feigenson & Halberda, 2011; Lyons, Ansari & Beilock, 2012; Mazzocco et al., 2011b).

Dietrich et al. (2015b) recommend accuracy as the most reliable and valid measure. The ANS tasks in the present research were paired-comparison tasks with quantities greater than 5 in Study 1, and greater than 10 in Study 3. Perceptual cues were controlled and performance was overall accuracy.

Summary. Counting and subitizing measures typically involve enumeration (i.e., identification) of nonsymbolic quantities (1 to 3 or 4 for subitizing; greater than 4 for counting), whereas most ANS tasks involve discrimination (i.e., comparison) of paired dot arrays. It is uncommon to see identification tasks in the large range or comparison tasks in the subitizable range. Some research has included identification of large quantities (e.g., Izard & Dehaene, 2008; Price, Clement & Wright, 2014), however performance on this task was not examined in relation to other ANS tasks, quantitative abilities, or math outcomes. Nonsymbolic comparison tasks in the 1 to 9 range are not uncommon, but rather than deriving separate subitizing and counting measures from this task, they are typically scored overall and labeled an index of the ANS. Soto-Calvo, Simmons, Willis, & Adams (2015) recent study was an exception; they employed a comparison task with quantities 1 to 3 as an index of subitizing. A recent study by Price and Fuchs (2016) separated nonsymbolic comparison into subitizable and nonsubitizable ranges for correlations—but then collapsed scoring over the range (1 to 9) for final analyses because they could not fit Weber fractions otherwise. The point here is that task range is typically confounded with task demand. ANS tasks require comparison, whereas subitizing and counting require identification. In the present research, identification and comparison tasks were completed in small, medium, and large ranges

Order judgment[1] tasks were also included. The ability to process ordinal information is something that is apparent in babies (Brannon, 2002) and monkeys (Brannon & Terrace, 1998). Ordinal tasks have been found to be dissociable from cardinality (identification) and discrimination (comparison) tasks, according to Nieder and Dehaene (2009). Lyons and Beilock (2011) found a very strong correlation of .70 between a symbolic order judgment task (digits 1-9) and adults' performance on a multi-digit arithmetic measure. In their study, order judgment completely mediated relations between ANS acuity and arithmetic performance. More recently, Lyons et al. (2014) found that symbolic ordering gained strength as a predictor of math outcomes across grades 1 to 6, becoming the strongest predictor at the end of their longitudinal study. The question of the importance of ordinal tasks merits further investigation. It is for this reason the present research included order judgment tasks, as well as identification and comparison tasks. Table 1.1 summarizes the domain-specific quantitative tasks included in the present research.

To summarize, organizing quantities into small, medium, and large ranges in both nonsymbolic and symbolic formats permitted a systematic examination of subitizing, counting, ANS, and symbolic quantitative skills in relation to each other, and in relation to mathematical outcomes in adults.

[1] This task was called "ordering" throughout this document, but participants did not actually put quantities/digits in order, rather they indicated whether the 3 stimuli were in ascending order or not.

Table 1.1.

Summary of Domain-Specific Quantitative Measures

	Format of Quantity					
	Symbolic			Nonsymbolic		
	Size of Quantity			Size of Quantity		
Task Demand	Small (1-3)	Medium (4-9)	Large (10-100)	Small (1-3)	Medium (4-9)	Large (10-100)
Identification: "Name the digit/quantity."	Naming of digits			Naming of quantities		
Comparison: "Which is more?"	Comparisons with digits			Comparisons with quantities		
Order Judgment: "Are these in order?"	Ordering with digits			Ordering with quantities		

Linking the Core Quantitative Systems to Symbols

In order to become proficient at symbolic mathematics, humans in literate cultures must learn number symbols (i.e., words and digits) and come to understand what these symbols represent—that is, numerical magnitude. Our constructed symbolic number system is a tool that extends the range of quantities we can conceptualize, while also adding precision to quantity tracking (e.g., tallies) and manipulation (e.g., arithmetic). But how do we acquire the symbolic number system? Researchers have posited that symbolic numbers are mapped to the existing core quantitative systems—that is, the subitizing and approximate number systems (e.g., Dehaene, 1997; Feigenson et al., 2004; Halberda et al., 2008; Leibovich & Ansari, 2016; Lyons & Ansari, 2015; Lyons & Beilock, 2011; Noël & Rousselle, 2011). However, there are different theoretical accounts of how these links develop, and in turn, how these core systems should relate to arithmetic performance. Three accounts and the associated research are described below. The *ANS-is-Key* account claims that individual differences in ANS acuity should predict math skill. In contrast, the *Subitizing-is-Key* account claims that individual differences in subitizing skills lead to math success. The *Subitizing+Counting* account places some importance on the subitizing system and even greater importance on the counting system as the key to math proficiency.

Account 1: ANS is Key. The *ANS-is-Key to mathematics* account claims that symbolic representations are grounded in the nonsymbolic magnitude representations provided by the ANS, a view put forth by various researchers (e.g., Dehaene, 1992, 1997; Dehaene & Cohen, 1995, 1997; Gallistel & Gelman, 1992). On this view, children already understand the rules and logic of the number system (e.g., one-to-one correspondence; cardinality) because these rules are inherent in the ANS (Gallistel & Gelman, 1992). Thus, the challenge that children face is in learning to map the counting sequence to their internal representations of quantity. It is argued that learning to count takes time to develop because it is not transparent or clear how the count words map to these internal approximate number representations. In sum, on this view, the approximate number system serves as the foundation for symbolic representations of quantity. According to this account, individual differences in ANS acuity should predict both symbolic number system knowledge, and later arithmetic skills; children must first master procedural counting skills to gain the tools necessary (e.g., symbols) to express what they inherently understand (Dehaene, 1997, Dehaene et al., 1998; Gallistel & Gelman, 1992; Gelman & Gallistel, 2004; Gilmore et al., 2010; Mazzocco et al., 2011a, 2011b; Piazza et al., 2010; Pica, et al., 2004; Wynn, 1990).

Since Halberda, Mazzocco and Feigenson's seminal paper (2008) in which they reported a correlation between ANS acuity in ninth grade and math performance in third grade, a great deal of research has investigated ANS acuity as a predictor of math performance. Recently Lyons and Ansari (2015) and Leibovich and Ansari (2016) reviewed the literature to determine whether symbolic quantitative abilities appear to be "grounded" in approximate number system representations; both reviews conclude that the evidence does not support this view.

The numerical distance or ratio effect observed in both nonsymbolic and symbolic quantity comparison has been taken as evidence that symbolic quantities are mapped to the nonsymbolic ones, and that the symbolic and nonsymbolic representations are quite similar (Leibovich & Ansari, 2016). However, research by Lyons, Ansari, and Beilock

(2012) in which participants made comparisons across symbolic and nonsymbolic formats suggests symbolic and nonsymbolic representations are not strongly connected. Lyons, Nuerk, and Ansari (2015) examined ratio effects within a large sample of children (Grades K to 6). They observed smaller ratio effects in symbolic relative to nonsymbolic comparison (30% versus 75%; i.e., indicating a lack of representational overlap), and the nonsymbolic ratio effects did not predict symbolic ones. These findings suggest that symbolic and ANS (nonsymbolic quantity) representations are not very similar, after all.

If symbols are grounded in the ANS, the ANS might scaffold the development of counting skill. Although the findings have been mixed, recent evidence from Odic, Le Corre & Halberda (2015) suggests that children can only map symbolic words to nonsymbolic quantities once they have acquired the cardinality principle, thus it seems the ANS is not the sole contributor to understanding how the count words (symbolic) represent quantity. According to the *ANS-is-key to mathematics* account, ANS skill would also be expected to predict later symbolic skills, but longitudinal research by Sasanguie, Defever, Maertens and Renvoet (2014) did not find a correlation between nonsymbolic processing in kindergarten and symbolic processing 6 months later. Fazio, Bailey, Thompson, and Siegler (2014) reviewed the literature and failed to find that symbolic quantitative skills mediated relations between nonsymbolic skills and math. There is even some evidence that—opposite of what is expected—symbolic quantity processing (age 3 to 4 years) predicts later nonsymbolic processing (Mussolin, Nys, Content & Leybaert, 2014). In contrast, Price and Fuchs (2016) found that relations between nonsymbolic comparison skills and numerical/mathematical outcomes were mediated by symbolic comparison skills. However, their nonsymbolic comparison task included the quantities 1 to 9, and thus (I would argue) is not a pure measures of ANS.

In support of the *ANS-is-key to mathematics* account, ANS acuity *is* correlated with mathematics achievement according to recent meta-analyses (r = .24, Schneider et al., 2016; see also Chen & Li, 2014; Fazio et al., 2014). However, ANS skills frequently fail to account for unique variance in math performance once symbolic quantitative skills are taken into consideration (Bartelet, Vaessen, Blomert, & Ansari, 2014; Brankaer, Ghesquière, & De Smedt, 2014; Castronovo & Göbel, 2012; Fazio et al. 2014; Fuhs & McNeil, 2013; Göbel, Watson, Lervåg, & Hulme, 2014; Holloway & Ansari, 2009; Kolkman, Kroesbergen, & Leseman, 2013; Lyons et al., 2014; Lyons & Beilock, 2011; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013; Szűcs, Devine, Soltesz, Nobes, & Gabriel, 2014; Toll & Van Luit, 2013; vanMarle, Chu, Li, & Geary, 2014).

To summarize, the evidence does not support the view that the ANS is the key to the development of counting or later symbolic skills (see Leibovich & Ansari, 2016). Furthermore, although ANS acuity is correlated with arithmetic performance, it tends to be an insignificant predictor once symbolic quantitative skills are taken into consideration. The ANS does not seem to be key to success in mathematics.

Account 2: Subitizing is Key. The *subitizing-is-key to mathematics account* suggests the counting system is scaffolded by the subitizing system, thus subitizing is key to the development of counting skills, later symbolic skills, and mathematical success. Butterworth and colleagues (e.g., Landerl et al., 2004) argue that basic numerical abilities are built on "early mechanisms for processing small numerosities" (p. 105). Thirty-five years ago, Starkey and Cooper (1980) suggested that the subitizing system acts as a precursor for verbal counting. More recently, Carey (2004) has been a proponent of this view. She proposes that children learn the arbitrary count string, and then map the meanings of the first count words to small quantities already represented in the subitizing system. Once children have gained some basic understanding of counting within this small set, they "bootstrap" their counting understanding to help them map the other count words to greater quantities. In time children are presumed to link their approximate representations of number (e.g., approximately 5) with exact quantities (e.g., exactly 5; consistent with the findings of Mussolin et al., 2014). Thus, proponents of this second view would argue that individual differences in subitizing would predict the acquisition of counting skill, and later math achievement in children, and perhaps in adults.

Subitizing and math achievement. Some evidence to support the claim that subitizing is crucial to math abilities comes from research with children with a specific math disability, known as dyscalculia. A few studies have found that children with dyscalculia have steeper subitizing latency slopes relative to their typically developing peers, suggesting that they were counting rather than subitizing in the "normal" subitizing range (Koontz & Berch, 1996; Landerl et al., 2004; Schleifer & Landerl, 2010).

Dot enumeration is also associated with math performance among typically-developing children. In a longitudinal study of children in Grades 2 to 9, children enumerated dots 1 to 9 (Reigosa-Crespo et al., 2011). Performance on this task was observed to be a significant unique predictor of arithmetic scores in Grades 2 to 9, with the exception of Grade 4 (see also Reeve, Reynolds, Humberstone, & Butterworth, 2012). Gray and Reeve (2016) recently found that dot enumeration performance (quantities 1 to 5) was associated with math ability profile membership. Children (ages 3- and 4-years) with better dot enumeration performance were (marginally) less likely to be assigned to the "poor math" profile, and more likely to be assigned to the "excellent math" or "good math, poor count sequence" profiles. These studies suggest that dot enumeration is an important predictor of math success up to the teenage years—however, the effects of subitizing and counting cannot be disentangled as performance in these studies collapsed across these two quantitative systems. Similarly, Sowinski et al. (2015) included subitizing, but it was combined with counting and symbolic magnitude comparison, rather than examined as a unique predictor of mathematical outcomes.

A few studies have isolated the effects of subitizing. LeFevre et al. (2010) found that subitizing was related to counting and was a unique predictor of concurrent nonverbal arithmetic (children age 4- to 6-years). Soto-Calvo et al. (2015) found that subitizing predicted cardinal counting; it was also correlated with numerical operations while controlling for age, but did not predict numerical operations uniquely in multiple regression analyses. Reigosa-Crespo et al. (2013) found that subitizing skill in third grade was uniquely predictive of mathematics fluency one year later; this was while controlling for domain-general skills like nonverbal reasoning and processing speed. Hannula-Sormunen, Lehtinen, and Rasanen recently found that subitizing performance at age 5 years predicted counting and spontaneous focusing on numerosity (SFON) at age 6 years (Hannula, Rasanen, & Lehtinen, 2007), which in turn predicted mathematical skills at age 12 years (Hannula-Sormunen, Lehtinen, & Rasanen, 2015). Subitizing exhibited a significant indirect effect on mathematical skills. These findings mostly held true when nonverbal IQ was added as a direct predictor of mathematical skill, with the exception that counting was no longer a significant direct predictor. Nonverbal IQ was not a significant direct predictor, either. Note that the sample-size was quite small ($N = 36$), so counting might have remained a significant predictor with a larger sample (i.e., if there was more power).

To summarize, there is some evidence that people with specific math disabilities demonstrate deficits in subitizing. Although some studies have not differentiated subitizing from counting skill, there is evidence that one or both of these processes are important, given the predictive strength of dot enumeration tasks. Some recent research has begun to link subitizing to counting and mathematical outcomes. So far the evidence supports the view that counting is scaffolded by the subitizing system—but more research is needed. The evidence does not really support the view that subitizing is the key to math success, however, as it was not a direct predictor in Hannula-Sormunen et al. (2015).

Account 3: Subitizing+Counting is Key. Like the second account, in the third account, the counting system is mapped onto the subitizing system—the difference is that the counting system is believed to be the key to math success (rather than subitizing); the ANS is presumed to become refined through counting. The *Subitizing+Counting-is-Key to mathematics* account was used to frame the current thesis. Like Butterworth (2010) and Carey (2001, 2004; Le Corre & Carey, 2008), Noël and Rousselle (2011) argue that the symbolic number system is first built on the subitizing system, because this system allows children to determine the exact quantity of small sets. Each object in the (small) set is an individual, thus allowing children to see that adding one to the set changes the value of the set. Children are thought to gradually learn how to count to four (this includes understanding one-to-one correspondence and cardinality), at which point they extend their understanding of counting with small numbers to counting larger numbers

(Noël & Rousselle, 2011; Carey 2001). Counting higher allows children to then determine the exact amount of quantities previously only represented in the approximate number system (Noël & Rousselle, 2011; Sarnecka & Carey, 2008). It is not until children have developed exact representations for numbers larger than four that they start to link these back to the ANS. Thus, Noël and Rousselle suggest that the links between ANS and the exact (counting) system explain why the acuity of the ANS gets better across development (Halberda & Feigenson, 2008) and posit that ANS deficits observed in dyscalculic children may be a *consequence* of an impaired counting and/or subitizing system, rather than a cause of the observed deficits in math. In sum, the *Subitizing+Counting-is-Key to mathematics* account highlights the importance of the counting system—which is built upon the subitizing system; thus, this account subsumes the *Subitizing-as-Key* account and places greater importance on counting.

Counting and Math Achievement. Consistent with Noël and Rousselle's (2011) account, many studies have linked early counting skills to math achievement in typically-developing children (Johansson, 2005; Locuniak & Jordan, 2008; Reigosa-Crespo et al., 2013; Stock, Desoete & Roeyers, 2009). For example, Aunola, Leskinen, Lerkkanen and Nurmi (2004) found that preschool counting skills predicted math skills in Grade 2. Geary and colleagues (Geary, 2011; Geary, Bow-Thomas & Yao, 1992; Geary, Hamson & Hoard, 2000; Geary, Hoard & Hamson, 1999; Geary, Hoard, Byrd-Craven, & DeSoto, 2004) and Landerl et al. (2004) have found that poor counting skills are related to math disabilities in children (see also Gersten, Jordan, & Flojo, 2005). Counting deficits have also been observed in children with low math achievement (Torbeyns, Verschaffel, & Ghesquiere, 2004). The research by Hannula-Sormunen et al. (2015) described previously is consistent with the *subitizing+counting* account, as counting was predicted by subitizing, and it uniquely predicted mathematical skill later on. Reigosa-Crespo (2013) also found counting to be a marginally significant predictor of mathematics fluency while controlling for domain-general processes. In a recent study by vanMarle et al. (2014) verbal counting and Give-N performance (a measure of understanding of the counting principle, *cardinality*) were found to uniquely predict mathematics achievement while controlling for demographic and domain-general cognitive skills. In sum, there is evidence that counting is important to the development of mathematical skills, however, little research has examined counting in relation to subitizing and ANS skill to determine their relative contributions to mathematical outcomes.

Summary. There are three different accounts of how core number systems become linked to the symbolic system and to arithmetic. The evidence is not in favour of the *ANS-is-Key* account. The evidence so far supports the *Subitizing+Counting-is-Key* over the *Subitizing-is-Key to mathematics* account, but further research is required. It remains unclear exactly how subitizing, counting, and ANS skill relate to one another and to math skill in adults. The common thread among these accounts is the importance placed on the core number systems and the importance placed on counting as the first link between the nonsymbolic core quantitative systems and the symbolic number system. As mentioned previously, I have adopted Noël and Rousselle's view (*Subitizing+Counting-is-Key* account), which highlights the importance of counting, the potential importance of subitizing, and places less emphasis on the importance of the ANS as predictors of math achievement.

Symbolic Versus Nonsymbolic Quantitative Skills

Even if the subitizing and/or ANS systems scaffold the development of the symbolic number system, do the nonsymbolic and symbolic quantity representations become “estranged”, as suggested by Lyons, Ansari and Beilock (2012)? Lyons et al. suggest that the links between nonsymbolic and symbolic representations are closely related for a time (e.g., in the primary school years) and then become somewhat independent of one another. The symbolic quantitative representations might supersede the nonsymbolic representations, thus partially or entirely reducing correlations among nonsymbolic skills and arithmetic skill observed earlier on in development. This de-coupling could occur because the relations among the digits (i.e., ordinal information) come to overshadow relations between the digits and the nonsymbolic quantitative representation (Nieder, 2009). Recent evidence from Lyons, Ansari and Beilock (2015) indicates that symbolic and nonsymbolic quantity processing do not share the same structures in the brain, suggesting that symbolic representations do not “inherit the representational structure of their nonsymbolic counterparts” (p. 484). Thus, consistent with Lyons et al.’s (2012) symbolic-estrangement claim, symbolic and nonsymbolic representations of quantity do not seem to be very similar or overlapping.

Furthermore, mounting evidence highlights the importance of symbolic quantitative skills. De Smedt et al. (2013) reviewed 25 studies, comparing the roles of nonsymbolic and symbolic number processing skills as predictors of math performance. Symbolic number processing skills tended to predict math performance (13 out of 17; 76%), while nonsymbolic number processing skills did not (11 out of 25; 44%). Schneider et al.’s recent meta-analysis (2016) identified a stronger effect of symbolic compared to nonsymbolic comparison skill when predicting mathematical outcomes (r = .30 versus .24); this difference was statistically significant. Vanbinst, Ansari, Ghesquiere & De Smedt (2016) recently compared the effect of symbolic processing on mathematics performance to the effect of phonological awareness on reading skill; they found these effects were the same size and concluded that symbolic processing is an important domain-specific predictor of arithmetic. Based on this pattern of results, Lyons and Ansari (2015) suggested future research focus on the ways to develop symbolic quantitative skills, since these skills directly predict math performance (see also Schneider et al., 2016).

The main goal of the present research was to identify domain-specific predictors of arithmetic performance. Given the evidence highlighting the importance of symbolic quantitative skills, it was therefore important to include both symbolic and nonsymbolic skills, in order to get a comprehensive account of unique predictors of mathematical outcomes. To summarize, organizing quantities into small, medium, and large ranges in both nonsymbolic and symbolic formats permitted a systematic examination of subitizing, counting, ANS, *and* symbolic quantitative skills in relation to each other, and in relation to mathematical outcomes in adults.

Domain-General Cognitive Skills

Domain-general cognitive skills have been found to predict mathematical outcomes (see references below), thus, in order to truly test the role of domain-specific quantitative abilities, domain-general cognitive skills were also taken into consideration. In Study 3 a broad range of domain-general cognitive skills were assessed, and combined to create a latent construct for use in structural equation modeling. The types of tasks included in the domain-general cognitive factor were quite consistent with the theoretical conceptualization of executive function, and thus the factor was labeled as such; see Cragg and Gilmore (2014) for a recent review on executive function in relation to mathematics (also, Raghubar, Barnes, & Hecht, 2010 and Peng, Namkung, Barnes, & Sun, 2015[2] for reviews *specific* to working memory). The executive function construct includes response inhibition (one's ability to inhibit information to effectively control responses), set shifting (one's ability to change tasks/change responses when faced with a new task), and working memory (Miller, Giesbrecht, Müller, McInerney, & Kerns, 2012; Miyake et al., 2000). The working memory component is typically consistent with Baddeley's (2001) conceptualization, in which the central executive manages attention and strategy selection, and controls the two slave systems (phonological loop and visual-spatial sketchpad). In the current research, five tasks tapped the following aspects of executive function: inhibition, attention, and working memory—both visual spatial, and phonological loop, as well as processing speed. The shifting component of executive function was not included (Cragg & Gilmore, 2014).

Processing speed is the speed with which basic cognitive operations are carried out (Salthouse, 1996). Speeded measures, such as speeded addition, depend not only on quantitative processes, but also on the speed with which one can access and respond to stimuli. Processing speed has been found to predict math achievement (Bull & Johnston, 1997; Chong & Siegel, 2008; Clark et al., 2014; Clark, Sheffield, Wiebe & Espy, 2013; Hecht, Torgesen, Wager & Rashotte, 2001). Inclusion of this measure was particularly

[2] "Measures that tap executive functions, such as inhibition, switching, or updating, were not considered to be WM measures in this meta-analysis, nor were simple span measures (storage without simultaneous processing) considered to tap WM." (p. 5, Peng et al., 2015)

important since response times serve as the performance measure for many of the included tasks.

Central executive components of working memory and executive functions have been found to positively relate to math outcomes (Clark et al., 2013; Clark et al., 2014; Friso-van den Bos, Kroesbergen, & van Luit, 2014; Hoard, Geary, Byrd-Craven, & Nugent, 2008; Kleemans, Segers, Verhoeven, 2011; LeFevre et al., 2013; Locuniak & Jordan, 2008; Mazzocco & Kover, 2007; Murphy, Mazzocco, Hanich, & Early, 2007; Noël, 2009; Passolunghi & Cornoldi, 2008; Passolunghi, Mammaralla, & Altoè, 2008; Swanson & Kim, 2007), such as arithmetic (Andersson, 2008), fact retrieval and accuracy (Marzocchi, Lucangeli, De Meo, Fini, & Cornoldi, 2002), number line task performance (Geary, Hoard, Nugent, & Byrd-Craven, 2008), and efficient strategy selection (Wu et al., 2008). Deficits in working memory have also been associated with impaired arithmetic fact fluency in children across grades 2 to 5 (Chong & Siegel, 2008) and with growth in fluency in children from grades 2 to 4 (LeFevre et al., 2013).

The visual-spatial component of working memory has been found to uniquely predict math achievement (Bull, Espy, & Wiebe, 2008; Hoard et al., 2008; Krajewski & Schneider, 2009; LeFevre et al., 2010; Rasmussen & Bisanz, 2005), especially in younger children, whereas the phonological loop predicts performance in older children (Andersson, 2008; Bull et al., 2008; Hoard et al., 2008; Noël, 2009; Rasmussen & Bisanz, 2005). To conclude, given the evidence that domain-general cognitive skills predict math achievement, a range of cognitive abilities were measured and used to create a domain-general cognitive factor for analyses; this factor was labeled executive function.

Experiential Factors

Age, education location, and perceived math competence were included as experiential factors in all three studies. Experiential factors were expected to predict arithmetic scores, but be unrelated to the core questions about the role of quantity processes in mathematical outcomes. Thus, controlling for these known predictors should improve the ability to detect differences related to quantity processes. A dichotomous Education Location variable was used to create two groups: Those educated in Asia and those educated elsewhere. This control variable was included because participants from Asian nations outperform those from other countries on international math assessments (Mullis, Martin, Foy, & Arora, 2012) and in research conducted at Carleton (e.g., LeFevre & Liu, 1997; Sowinski, Dunbar, & LeFevre, 2014), probably due to a stronger emphasis on math proficiency within those cultures. Thus, the presence of Asian-educated individuals was expected to uniquely predict mathematical outcomes. Perceived math competence (PMC) was also included; participants self-reported their perceived skill in mathematics, as well as nervousness about, and avoidance of, mathematical situations. Math anxiety has been shown to have negative effects on mathematics performance; see Suárez-Pellicioni, Núñez-Peña, and Colomé (2016) for a recent review. Some researchers theorize that math anxiety should relate to symbolic or nonsymbolic processing (e.g., magnitude comparison; Ashcraft, Krause, Hopko, Berch, & Mazzocco, 2007; Dietrich, Huber, Moeller, & Klein, 2015a; Maloney, & Beilock, 2012). PMC is thought to reflect participants' mathematical experiences. Feeling more competent in and less nervous/avoidant towards math was expected to positively relate to arithmetic fluency, accounting for unique variance beyond quantitative skills (Ashcraft, 2002).

Current Research

The overarching goal of the current research was to systematically examine (nonsymbolic) subitizing, counting, and approximate number system skills, as well as symbolic quantitative skills in relation to one another and in relation to mathematical outcomes. Measures of domain-general cognitive abilities were also collected, and will be examined in relation to domain-specific cognitive skills. The six specific goals or questions that were addressed in this thesis are summarized here. Within each section, goals are described followed by how they were addressed in the present research. Table 1.2 provides a summary of research questions and goals by study. Specific hypotheses are also discussed and numbered below (e.g., H1, H2, etc.), and summarized in Table 1.3.

Goal 1: To examine the relation between subitizing and mathematical performance in adults. Despite the fact that the subitizing and approximate number systems have been identified as core number systems (Feigenson et al., 2004)—and thus are both candidates for symbol grounding (Leibovich & Ansari, 2016; Lyons & Ansari, 2015), very little research has looked at subitizing in relation to mathematical skill. Only a small number of studies have examined subitizing in relation to number and math skill development among children (Gray & Reeve, 2014, 2016; Landerl et al., 2004; LeFevre et al., 2010; Schleifer & Landerl, 2010; Soto-Calvo et al., 2015) and none among adults. To address this issue, the present research examined subitizing in relation to: a) math skill (Studies 1, 2, & 3), b) other domain-specific skills (e.g., ANS; Studies 1, 2, & 3) and c) domain-general cognitive skills (Study 3). Subitizing and ANS were expected to be uncorrelated as they are thought to be distinct quantitative systems (H1). When subitizing and ANS were the only two quantitative predictors, subitizing was expected to predict mathematical outcomes, while ANS was not (H2), given the evidence that ANS is not key to mathematical development. Subitizing skill was hypothesized to predict counting skill (H3) based on the Subitizing-is-Key and Subitizing+Counting-is-Key accounts described previously. Subitizing was also expected to predict symbolic identification and comparison skills (H4) based on the assumption that symbolic representations are scaffolded by the quantitative representations inherent in the subitizing system. Relations among subitizing and mathematical outcomes are expected to be mediated by symbolic quantitative skills (falls under H11).

Goal 2: To clearly define and measure the Approximate Number System. A large body of recent research has examined the approximate number system in relation to math skill among adults (see Dietrich et al., 2015b for methodological recommendations). Among this research, one issue is the variation in quantities used to index the ANS. Some studies include magnitude comparison tasks with the quantities 1 to 9 as a measure of the ANS (e.g., Lyons & Beilock, 2011; Mundy & Gilmore, 2009). The approximate number system is defined as a system for processing large numerosities, and I argue that quantities 1 to 9 are not large enough to unequivocably index the ANS (Dietrich et al., 2015b also make this claim). Tasks that include quantities between 1 and 9 may actually be indexing other systems, like subitizing or counting. At the very least, ANS measures should not include quantities 1 to 4 (Dietrich et al., 2015b) as these quantities are apprehended quickly and *exactly*—not approximately (Mandler & Shebo, 1982; Trick & Pylyshyn, 1994). To address this issue, the nonsymbolic magnitude comparison task indexing ANS acuity in Study 1 (the Panamath; Halberda, Mazzocco et al., 2008) had a minimum quantity of 5, and thus did not tap subitizing skill. In Study 3, the subitizing, counting, and ANS systems were clearly demarcated by specifying small (1 to 3) medium (5 to 9), and large (10 to 100) quantity ranges; nonsymbolic tasks within these ranges were expected to tap subitizing, counting, and ANS respectively.

As described above, researchers have also identified potential issues with ANS stimuli and scoring methods (see Dietrich et al., 2015b for a review). Of note, efforts should be taken to ensure that ANS magnitude comparison tasks (also known as quantity discrimination tasks; e.g., “Which set of dots has more?”) control for visual cues that correlate with numerosity. These cues include overall area, array perimeter, and size of individual dots, which should be varied within the task to prevent participants from relying on non-numeric perceptual cues to aid performance (Gebuis & Reynvoet, 2011). That is, by varying the ANS stimuli along these dimensions, participants must focus on the numerosities in order to be successful. To address this issue, comparison stimuli were designed so that area positively correlated with the numerosity for half of the trials; area was negatively correlated with numerosity for the other half. Dots also appeared in various sizes within each pair. Scoring methods for magnitude comparison tasks should also be taken into consideration; overall accuracy demonstrates superior test-retest reliability in comparison to the Weber fraction and numerical distance effect (Inglis & Gilmore, 2014). Thus, accuracy was the ANS performance measure in Studies 1 and 3.

As described above, evidence is mounting to indicate that ANS skill is weakly related to mathematical outcomes, and frequently fails to uniquely predict mathematical outcomes when symbolic skills are also considered (Schneider et al., 2016). Therefore, ANS skill was not expected to predict mathematical tasks requiring exact computation, such as arithmetic (H5). I can see no logical reason why an approximate quantity representation (the ANS) would be employed when exact calculation is required—especially when one has an exact quantity representation to rely on (i.e., the symbolic number system). However, it seems reasonable for the ANS to be employed when a numerical task requires approximate rather than exact processing—such as a number line estimation task. Thus, ANS skill was hypothesized to predict the symbolic and nonsymbolic number-to-position number line tasks included in Study 3 (H6).

Goal 3: To clearly define counting and examine it as a predictor of mathematical performance in adults. In the Subitizing+Counting account put forth by Carey (2004), counting is the crucial skill required to become proficient in math. Although researchers have examined the development of counting skills among young children and how these skills relate to math development, there is little research on counting in relation to math achievement among adults. Some research has included symbolic and nonsymbolic magnitude comparison in the 1 to 9 range—which includes both the small and medium ranges as I have defined them. The evidence with babies and monkeys (Feigenson, et al., 2002; Hauser et al., 2000; Xu & Spelke, 2000), suggests a gap between the subitizing quantities (1 to 4) and quantities large enough to employ the ANS, suggesting there is a need for a clearly defined intermediate range. In the present research, the medium range was defined as quantities 5 to 9, and nonsymbolic tasks in this range were assumed to tap counting skill. Based on the Subitizing+Counting-is-Key account, counting was hypothesized to predict symbolic identification and comparison tasks (H7), and was also examined as a direct predictor of mathematical outcomes. In Study 3, nonsymbolic quantitative skills in the medium range were expected to predict symbolic quantity identification and comparison.

Goal 4: To see if different kinds of tasks index the subitizing, counting, and approximate number systems. As described above, different tasks have typically been used to index subitizing and counting versus ANS skill. Subitizing and counting tasks tend to be identification tasks in which participants are asked to state the number of dots or objects shown (e.g., Gray & Reeve, 2014, 2016).[3] ANS tasks have tended to be simultaneous magnitude comparison (Dietrich et al., 2015b), however some recent studies have employed nonsymbolic order judgment tasks (e.g., Lyons & Beilock, 2009, 2011). The literature lacks nonsymbolic identification tasks in the large (ANS) range, and comparison tasks in the small (subitizing) range, thus the present research included such tasks. Doing so permits an examination of task by range in relation to math performance. It was impractical to limit order judgment tasks (nonsymbolic and symbolic) to subitizing and counting ranges, however, given the small number of possible combinations of three quantities in this range (1, 2, and 3 only). Instead, ordering tasks were separated into small-medium (1 to 9) and large (10 to 100) ranges. The goal here was to separate the ANS from the subitizing/counting systems in the nonsymbolic ordering tasks. It was hypothesized that nonsymbolic identification, comparison, and ordering tasks from the same range (small, medium, and large) would tap the same underlying constructs—namely, subitizing, counting, and ANS skills, respectively (H8). A similar approach was employed with the symbolic tasks (grouping them together by range), but in contrast to the nonsymbolic tasks, they were assumed to tap the one underlying construct—namely symbolic quantity processing. Symbolic tasks were expected to form a hierarchy by task

[3] The research with babies has been a bit more varied, including habituation, active search (e.g., how long spent looking for hidden objects once one or more have been removed), and selection preference tasks (e.g., crawling to select one of two containers of crackers).

complexity. Identification was presumed to be the least complex, followed by comparison, followed by ordering. It was hypothesized that identification would predict comparison, and comparison would predict ordering (H9).

Goal 5: To contrast symbolic versus nonsymbolic quantitative tasks as predictors of mathematical outcomes. There is mounting evidence that symbolic magnitude comparison performance uniquely predicts math performance while nonsymbolic comparison does not (Schneider et al., 2016). By comparison, relatively little research has included symbolic and nonsymbolic versions of identification and ordering tasks as predictors of mathematics outcomes. For this reason, the present research included the same identification, comparison, and ordering tasks in both symbolic and nonsymbolic formats. Thus, the present research examined the relative contribution of symbolic and nonsymbolic quantitative skills to mathematical outcomes. It was hypothesized that nonsymbolic quantitative skills would not uniquely predict mathematical outcomes once symbolic quantitative skills were taken into consideration (H10). Given the anticipated symbolic task hierarchy described in the previous section, symbolic ordering was hypothesized to be the best predictor of mathematical outcomes in Study 3 (H11).

Goal 6: To explore the role of domain-general cognitive abilities as predictors of mathematical outcomes. As described previously, past research has demonstrated that domain-general cognitive abilities—like executive function and working memory—predict math performance, therefore, it was important to consider these domain-general skills in the present research. Doing so permits an examination of executive function in relation to subitizing, counting, ANS, and symbolic quantitative skills (which has probably not been done in past research, as these quantitative systems have not all been examined concurrently). Domain-general cognitive skills were expected to predict domain-specific skills, and indirectly predict mathematical outcomes (H12). Experiential variables were also expected to uniquely predict mathematical outcomes (H13). The present research included education location, age, and perceived math competence as indexes of experience.

Table 1.2.

Summary of Study Goals

Goal	How Goal Was Addressed	Studies
1. To examine the relation between subitizing and mathematical performance in adults	• Small (subitizing) range defined as 1 to 3 • Subitizing examined in relation to: o math skill o other quantitative skills o domain-general cognitive abilities	1, 2, 3 1, 2, 3 1, 2, 3 3
2. To clearly define and measure the Approximate Number System	• Large (ANS) range defined as 10 to 100 • ANS measures exclude 1 to 4 (subitizing range) • Nonsymbolic magnitude comparison tasks o Perceptual cues controlled (e.g., area, dot size) o Accuracy used (not NDE or Weber fraction)	3 1 1 & 3 1 & 3
3. To clearly define counting and examine it as a predictor of mathematical performance in adults	• Medium (counting) range defined as 5 to 9 o Distinct from Small (subitizing; 1 to 3) and Large (ANS; 10 to 100) ranges • Counting examined in relation to: o math skill o other quantitative skills o domain-general cognitive abilities	3 3
4. To see if different kinds of tasks index the subitizing, counting, and approximate number systems.	• Identification & comparison tasks in small (1 to 3), medium (5 to 9), and large (10 to 100) ranges • Ordering in small-medium (1 to 9) and large (10 to 100) ranges	3 3
5. To contrast symbolic versus nonsymbolic quantitative skills as predictors of mathematical outcomes	• Symbolic and nonsymbolic versions of identification, comparison, and ordering tasks in small, medium, and large ranges	3
6. To explore the role of domain-general cognitive abilities as predictors of mathematical outcomes	• Multiple domain-general skills measured • Domain-specific skills expected to mediate relations between domain-general skills and math performance	3

Table 1.3.

Specific Research Hypotheses

Category	Hypotheses	Study
Nonsymbolic small (1 to 3): **Subitizing**	1. Subitizing skill will not relate to ANS skill	1 & 3
	2. Subitizing will predict mathematical outcomes but ANS will not	1
	3. Subitizing skill will predict counting skill	3
	4. Subitizing skill will predict symbolic identification and comparison	2 & 3
Nonsymbolic large: **ANS**	5. ANS skill will not uniquely predict *exact* mathematical calculation (e.g., arithmetic)	1 & 3
	6. ANS skill will predict approximate numerical outcomes, i.e., number line performance	3
Nonsymbolic medium (5 to 9): **Counting**	7. Counting will predict symbolic identification and comparison	3
Task & format	8. Nonsymbolic identification, comparison, and ordering tasks from the same range (small, medium, and large) will tap the same underlying constructs (subitizing, counting, and ANS systems)	3
	9. Less complex symbolic quantitative skills will predict more complex ones (i.e., identification will predict comparison, comparison will predict ordering)	2 & 3
	10. Nonsymbolic quantitative skills will not predict mathematical beyond symbolic quantitative skills	2 & 3
	11. Symbolic ordering will be the best quantitative predictor of exact mathematical outcomes	3
Domain-general cognitive skills	12. Domain-general cognitive skills will predict domain-specific quantitative skills and have indirect effects on mathematical outcomes	3
Experience	13. Experiential factors will uniquely predict mathematical outcomes	1, 2, 3

CHAPTER 2: STUDIES 1 AND 2

In the present chapter I examined subitizing as a predictor of arithmetic fluency. As described in Chapter 1, both the approximate number (ANS) and subitizing systems are presumed to be core nonsymbolic quantitative systems that may support the development of symbolic representations of quantity (Feigenson et al., 2004). A considerable amount of research has examined ANS skill in relation to math skills in both children and adults (Schneider et al., 2016), whereas relatively little research has examined the role of subitizing—especially among adults. Hence, in the present research subitizing was examined both in relation to other quantitative skills, and in relation to arithmetic fluency among adults. If subitizing is found to uniquely predict mathematical outcomes in adults, this evidence would support the view that subitizing is a core quantitative skill that forms the basis for children's acquisition of the symbolic number system.

Typical subitizing tasks require participants to enumerate small nonsymbolic quantities (i.e., 1 to 3 or 4 dots) intermixed with larger quantities (i.e., 5 to 9). Sometimes researchers calculate inverse efficiency by dividing the overall RT by overall accuracy to determine overall task efficiency (e.g., Reigosa-Crespo et al., 2012), however this performance measure does not isolate subitizing performance from counting performance (see also Gray & Reeve, 2016). To assess individual differences in subitizing speed, researchers can examine four RT parameters (as outlined by Grey & Reeve, 2014): 1) the subitizing range (i.e., determined by the point where discontinuity occurs in the RT slope), 2) the subitizing RT intercept, 3) the subitizing RT slope, and 4) the counting RT slope. Reeve and colleagues (Grey & Reeve, 2014; Reeve et al., 2012) have even assigned children to overarching subitizing profiles based on performance on these four parameters (grouping like-performing children together).

Here I argue that typical measures of subitizing pose certain challenges for implementation and interpretation. Administering them requires a computer for presentation of stimuli and the collection of RTs. Relative to calculating inverse efficiency, obtaining the four subitizing RT parameters described above requires substantial data processing. Furthermore, creating latent profiles based on RT parameters requires knowledge and understanding of that technique—and may not result in predictive validity. For example, although Gray and Reeve (2014) extracted RT parameters and created (subitizing) latent profiles, ultimately inverse efficiency seemed to be a better predictor of arithmetic performance than the children's subitizing profile. Although not entered as simultaneous predictors in the same model, inverse efficiency had a greater standardized beta coefficient (-.43) than profile-membership (.25). Inverse efficiency was also a significant predictor of subtraction while profile-membership was not.

The finding that inverse efficiency, which collapses across a large range of set sizes, is a better predictor of math performance than subitizing profile may indicate that counting, rather than subitizing, predicts arithmetic performance. It could also be the case that subitizing did not predict arithmetic performance because the latent profiles did not sufficiently capture (or summarize) individual differences in subitizing skill. It is my view that a subitizing measure that a) specifically taps the subitizing range (not the counting range), b) is easy to administer and score, c) results in a continuous variable, and, d) includes a measure of the *speed* (thus capturing at least some of the same information as RTs) would provide a useful alternative to previously-employed subitizing indicators (i.e., RT parameters).

To avoid the potential threats to construct validity and to simplify and standardize the measurement process, in the present research subitizing performance was measured with a novel paper-and-pencil task that was quick and easy to administer (less than two minutes) relative to computerized enumeration tasks. Participants simply identified (i.e., named) small nonsymbolic quantities—that is 1 to 3 dots—as quickly as possible while being timed. The quantities 1 to 3 were the only ones included, and thus—compared to inverse efficiency—this task may be a purer measure of subitizing. The score on the task is a speed score, specifically calculated as items-per-second (corrected for errors, if any). The inclusion of two forms (pages) permits the calculation of parallel-form reliability.

In Study 1, the subitizing and approximate number systems were examined as predictors of arithmetic fluency. Speeded naming of letters was also included to control for naming speed, thus isolating the quantitative aspect from the speeded aspect of the subitizing measure. To my knowledge, Study 1 is the first to a) include subitizing as a predictor of arithmetic skill among adults, and b) examine subitizing and ANS skills simultaneously in relation to each other and in relation to arithmetic performance. Given the evidence that the subitizing and approximate number systems are distinct nonsymbolic numerical systems, they were not expected to be correlated. Subitizing was hypothesized to predict arithmetic performance, presuming that the correlation observed in childhood persists into adulthood. Given the recent meta-analysis that showed that the average correlation between ANS acuity and math outcomes was .241 (Schneider et al., 2016), ANS acuity was expected to correlate with arithmetic fluency. ANS acuity was not expected to predict arithmetic fluency beyond subitizing.

Study 2 included the subitizing task (rapid identification of small quantities), as well as an analogous rapid identification of digits task as the naming control. Study 2 also included symbolic and nonsymbolic quantity comparison in the small-medium range (1 to 9 dots/digits). The four quantitative tasks were examined in relation to each other, as well as in relation to arithmetic performance. According the Subitizing-is-Key and Subitizing+Counting-is-Key accounts, the subitizing system supports the development of the verbal counting system, and in turn the acquisition of the symbolic number system, thus, in Study 2, subitizing was expected to predict the other quantitative tasks. Nonsymbolic comparison (which in the medium range is expected to tap subitizing and counting skill) was also expected to predict symbolic comparison. Symbolic identification is expected to precede the symbolic comparison abilities, thus rapid digit naming was expected to predict symbolic comparison as well.

The inclusion of the same tasks in both symbolic and nonsymbolic formats also permits a test of the hypothesis that nonsymbolic quantitative skills do not predict arithmetic fluency beyond symbolic quantitative skills. This hypothesis is based on the evidence that nonsymbolic magnitude comparison (indexing the ANS) does not uniquely predict mathematical outcomes when symbolic skills are also taken into consideration (see De Smedt et al., 2013). All quantitative measures (subitizing speed, digit naming speed, and symbolic and nonsymbolic quantity comparison) were expected to correlate with arithmetic fluency, but only the symbolic comparison and digit naming performance were expected to uniquely predict arithmetic fluency in path analyses.

In both studies, relations among basic quantitative skills and arithmetic performance were examined while controlling for experiential factors—that is age, location in which participants received their early mathematical education, and perceived math competence (PMC). As described in Chapter 1, experiential factors were expected to contribute to overall variability in arithmetic scores, but to be unrelated to the core questions about the role of quantity processes in calculation performance. Thus, these known predictors were controlled for in analyses predicting mathematical outcomes.

Study 1

Hypotheses

Study 1 tested three of the hypotheses summarized in Table 1.3.

- H1: Subitizing skill will not relate to ANS skill.
- H2: Subitizing but not ANS will predict arithmetic fluency
- H13: Experiential factors will uniquely predict arithmetic fluency

Method

Participants. Carleton University students (*N* = 167) participated in exchange for class credit or payment. Participants ranged in age from 17 to 60 years, with a median age of 20 years (60% female; 96% undergraduates). Thirteen participants did not report their age, however age was estimated in path analyses. Approximately 68% of participants completed their high school education in Canada; 16% completed high school in China or Hong Kong, and the remaining 16% completed high school in other countries. Most participants (66%) reported English as their first language. Other first languages were Chinese (18%), Arabic (4%), French (3%), and Spanish (2%), as well as a range of languages reported by only one or two participants (7% in total).

Procedure. The tasks included in the present study were secondary to one of four primary experiments on numerical cognition (Bourassa, 2014; Bouskill, 2013; Curtis, 2012; or Schildknecht, 2012). Participants completed measures of subitizing speed, approximate number system acuity (measured with the Panamath), arithmetic fluency, and letter naming, as well as a questionnaire to gather demographic information and measure perceived math competence (PMC). The tasks are described in the Measures section below.

The order of the tasks within each experiment was not constrained, but was decided by the individual experimenters. Sixty of the participants in the current study were participants in Bourassa's Master of Arts thesis (2014) which focused on ordinal judgments in relation to arithmetic fluency. Bourassa did not report the order of the tasks, so order of task administration is unknown. Thirty of the participants in the present study came from Bouskill's independent research project (2013) which focused on eye-movement patterns while solving subtraction problems. Bouskill administered the measures used in the current analysis after the main experiment; first the arithmetic fluency measure, followed by the subitizing and letter naming tasks (order unknown), the Panamath (ANS measure), and finally the questionnaire (demographic information and perceived math competence). Thirty-eight participants came from Curtis' Master of Cognitive Science thesis (2012), which examined the switch costs associated with addition and multiplication via eye-tracking data. Participants completed the eye-tracking data collection first, then the other measures (subitizing, letter naming, Panamath, arithmetic fluency; order unknown), ending with the questionnaire. Forty of the participants came from Schildknecht's independent research project (2012), which focused on subtraction problems recast as addition with negative numbers (e.g., -7 + 2). Schildknecht administered the tasks in the following order: Subtraction experiment, subitizing, letter naming, arithmetic fluency, Panamath, and questionnaire. Thus, the majority of participants completed the various measures used in the present research after they had solved arithmetic problems, and the questionnaire was typically completed last, suggesting that the perceived math competence measure was reflective of on-line experiences and thus well-informed.

Measures.

Demographic variables. A dichotomous control variable called *education location* was created to indicate whether participants were educated in Asia (2) or elsewhere (1). Age in years was also included as a predictor of arithmetic fluency.

Perceived math competence. Along with demographic information, participants self-reported their perceived math competence (PMC). Descriptive statistics and item correlations are shown in Table 2.1. These items were entered into a principle components analysis, to create a single PMC variable for further analyses. A single component was extracted and accounted for 66% of the variance in these measures. Component loadings are also shown in Table 2.1.

Letting naming: Rapid identification of letters. Participants named the letters "A", "M", and "C" as quickly as possible. Letters were presented in four rows with six stimuli per row, for a total of 24 stimuli per page. The three letters were presented randomly. The experimenter recorded the time taken to name all letters on the page as well as errors. An items-per-second score was calculated for each form using the following formula: Items-per-second = ((number of items)* – number of errors on form)/time on form. The correlation between the two forms (Form A and Form B) was $r(167) = .87$, indicating good parallel-forms reliability. The two scores were averaged to create a single variable for use in path analyses.

Table 2.1.

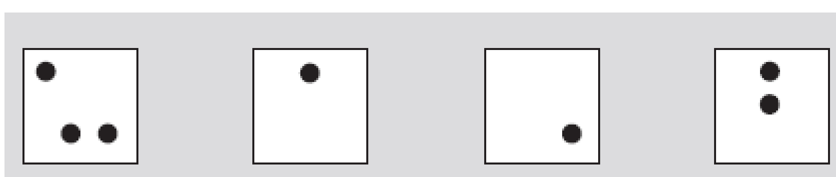
Perceived Math Competence Descriptive Statistics and PCA Results

Items	Mean		*SD*		Comm.		Component Loading		Item Correlations				
									1.	2.	3.	4.	5.
	S1	S2	S1	S2	S1	S2	S1	S2					
1. Please rate your level of basic mathematical skill (e.g., skill at arithmetic).	5.0	5.4	1.3	1.4	.58	.41	.76	.64	--	.54	.47	.22	.31
2. Please rate your level of mathematical skill in more complex areas of mathematics (e.g., calculus, algebra).	4.3	4.0	1.6	1.7	.72	.62	.85	.78	.64	--	.59	.44	.44
3. How often do you avoid situations involving mathematics?	4.4	4.1	1.8	1.9	.72	.74	.85	.86	.51	.64	--	.67	.50
4. How often do you find that situations involving mathematics make you nervous?	4.5	4.4	1.8	1.8	.67	.60	.82	.77	.50	.57	.68	--	.57
5. How difficult was mathematics for you in high school?	4.4	4.4	1.8	1.8	.59	.55	.77	.74	.45	.57	.58	.53	--

Note: $N = 167$ for Study 1 (S1) and $N = 89$ for Study 2 (S2). Item correlations are provided below the diagonal for Study 1 and above for Study 2. Comm. = Communality. For all items, participants responded with the numbers from 1 to 7. For Items 1 and 2, response options were 1 = *Very Low*, 4 = *Moderate*, and 7 = *Very High*. Response options for Items 3 and 4 were 1 = *Almost Always*, 4 = *Sometimes*, and 7 = *Almost Never*. For Item 5, response options were 1 = Extremely Difficult, 4 = Moderately Difficult, and 7 = *Not at All Difficult*. Responses in between anchors were not labeled.

Subitizing: Nonsymbolic rapid identification of small quantities (1 to 3). The subitizing task was designed, administered, and scored in the same way as the rapid identification of letters task. Participants rapidly identified small nonsymbolic quantities (1 to 3 dots) as the measure of subitizing speed. As shown in Figure 2.1, each dot stimulus was presented within a box to clearly distinguish one set of dots from the next. The locations of the dots varied across items (i.e., they were not all arranged in patterns, as on dice). There were four rows of six boxes, for a total of 24 stimuli per form; there were two forms (A and B). The time to name the stimuli on each form was recorded, as were errors (which were rare). Items-per-second scores were calculated for each form using the same formula used for the rapid identification of letters (shown previously). The inclusion of two forms permitted parallel-forms reliability analyses; the observed correlation between the two forms was $r(167) = .86$, indicating good parallel-forms reliability. The two scores were averaged to create a single subitizing variable for use in path analyses.

Figure 2.1. Example of subitizing stimuli.



Approximate number system (ANS): Nonsymbolic comparison of large quantities. ANS system acuity was assessed with the Panamath task (www.panamath.org; Halberda et al., 2008) via computer. Participants completed 256 nonsymbolic quantity comparisons in the large range (5 to 21 dots). Each trial consisted of a simultaneous paired-comparison of two sets of dots; yellow dots appeared on the right side of the screen and blue on the left. Participants indicated which set was greater (i.e., had more dots); the yellow set was selected by pressing the letter "J", and blue by pressing the letter "F" on the keyboard. The dots were presented until the participant responded or for a maximum of 600ms. Each trial was followed by a 200ms backward perceptual mask (i.e., a pixelated yellow and blue rectangle), and a fixation cross appeared between trials (a "+" sign in the middle of the screen). Participants pressed the space bar to initiate each trial. No feedback was provided.

Dots appeared in various sizes; the average dot size was 36 pixels with a 20% maximum variation in size. In half of trials, average dot size of the two sets was equated, resulting in the larger quantity occupying more area than the smaller quantity. For the other half of trials, the area of the two quantities was equated (i.e., the larger quantity had smaller dots on average). Varying the area and the dot size prevented participants from performing well on the tasks based on dot size or area alone—thus encouraging them to focus on numerosity in order to be successful. The ratio information for the stimuli, as provided by the Panamath creators, is presented in the Table 2.2. Accuracy (percentage correct) on the Panamath served as the measure of ANS acuity. Two participants had extreme accuracy scores, raising concerns about the validity of their performance; these scores were removed from further analyses.

Table 2.2.

Ratio Information for the Panamath (ANS) Task

Proportion Bins	Proportions as Decimals	Corresponding Ratios	Percentage of Trials From Bin
1.10 to 1.19	0.90 to .084	19:21 to 16:19	25%
1.19 to 1.29	0.84 to 0.78	16:19 to 7:9	25%
1.33 to 1.44	0.75 to 0.69	6:8 to 9:13	25%
2.29 to 2.48	0.44 to 0.40	7:16 to 7:17	25%

Arithmetic fluency. Arithmetic fluency was the dependent variable in Study 1; it was measured with the Calculation Fluency Test (CFT; Sowinski et al., 2014). The CFT is a speeded multi-digit arithmetic task. Participants were given one minute for each subtest; the three subtests were addition, subtraction, and multiplication (one page per operation). Performance was total correct (summed across the three subtests).

Results

Descriptive Statistics. Descriptive statistics for the variables are presented in Table 2.3. Note that for the letter naming and subitizing measures, descriptive statistics are presented for both raw and summary scores. Raw scores are presented to give an indication of the length of time required to administer the tasks.

Correlations. Correlations among variables are shown in Table 2.4. Age was weakly correlated with arithmetic fluency; performance increased with age. Education location was strongly correlated with arithmetic performance. That is, those educated in Asia had higher arithmetic fluency scores relative to those not educated in Asia. The perceived math competence (PMC) component was also moderately correlated with arithmetic fluency, as was letting naming. Letter naming was highly correlated with subitizing, presumably due overlapping task structure and requirement for speeded responding. The high correlation between letter naming and subitizing implies that including both as predictors of arithmetic fluency will be a particularly stringent test of subitizing skill's predictive strength. As hypothesized, subitizing speed was not significantly correlated with ANS accuracy, supporting the hypothesis these are separate quantitative systems (H1). Arithmetic fluency was moderately correlated with subitizing speed and weakly correlated with ANS accuracy.

Table 2.3.

Study 1 Descriptive Statistics for Observed Variables

	Raw Scores				Summary Scores					
	Rapid Identification				Rapid Identification Items-per-second		ANS Accuracy	Arithmetic Fluency (CFT)		
	Letters		Subitizing		Letters	Subitizing		Add.	Subtract.	Multipl.
	Time	Err.	Time	Err.						
Min	4.4	0	5.6	0	1.8	1.6	64.8	2	0	0
Max	14.1	3	18.5	1	5.3	4.1	93.0	35	33	33
Mean	7.5	0.13	9.5	0.10	3.3	2.7	83.4	16.7	12.6	8.2
SD	1.6	0.37	2.2	0.30	0.65	0.48	4.75	6.58	5.90	5.88

Note. $N = 167$. Err. = number of errors. ANS = approximate number system. NS Comp lg= Nonsymbolic comparison of large quantities. Calculation Fluency Test (CFT). Add. = Addition subscale. Subtract. = subtraction subscale. Multipl. = multiplication subscale. Time is measured in seconds. CFT scores are total correct. Descriptive statistics for PMC (Perceived Math Competence) variables are presented in Table 2.1.

Table 2.4.

Study 1 Correlations Among Predictors and Arithmetic Fluency

	1.	2.	3.	4.	5.	6.
1. Age	--					
2. Education location	.08	--				
3. Perceived math competence	-.04	.09	--			
4. Letter naming	.04	.03	**.18***	--		
5. Subitizing	-.05	.09	.14	**.71****	--	
6. ANS accuracy	.04	.10	**.24****	-.04	.09	--
7. Arithmetic fluency	**.19***	**.60****	**.31****	**.22****	**.30****	**.16***

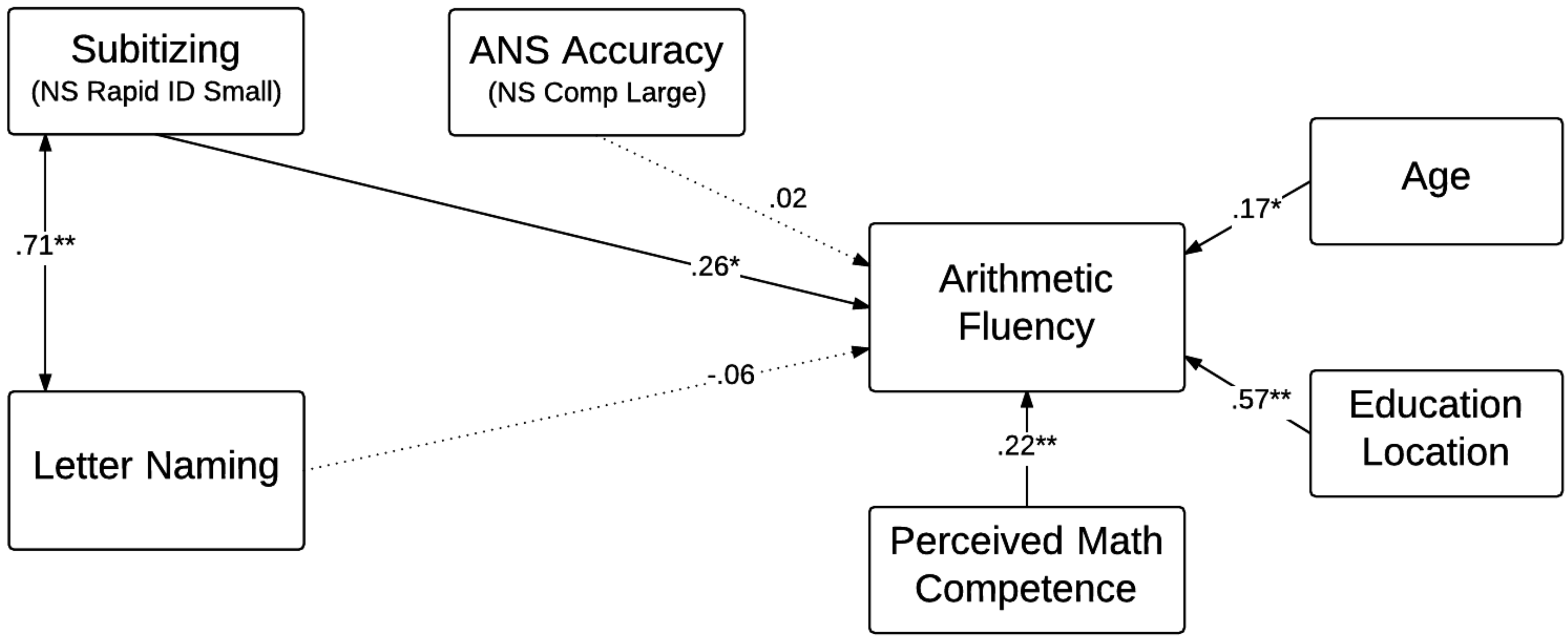
Note: $p \leq .05$, ** $p \leq .01$. $N = 167$. Education location was coded 0 = Educated elsewhere, 1 = Asian-educated. ANS accuracy is percent correct.

Path analysis. A path analysis was conducted with the statistical program MPlus (Muthén & Muthén, 1998-2012) with the full information maximum likelihood estimation. Subitizing speed, ANS accuracy, letter naming, age, education location, and perceived math competence were set to predict arithmetic fluency. Letter naming was set to correlate with subitizing. A diagram of the path model is shown in Figure 2.2. Single-headed arrows represent regression paths; the values on these arrows represent standardized regression coefficients. Double-headed arrows indicate correlations.

The fit indices examined to determine model and their cut-offs are described here (taken from Hu & Bentler, 1999; Schreiber, Nora, Stage, Barlow & King, 2006). The Chi-square test of model fit (χ^2) is ideally not significant, indicating that the model specified fits and does not differ significantly from the observed data. The root mean square error of approximation (RMSEA) should be less than .06 to .08; the Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) should equal or exceed .95. The standardized root mean square residual (SRMR) should be less than or equal to .08.

The observed model fit statistics were as follows: $\chi^2(8) = 15.8$, $p = .046$; RMSEA = .08 (90% CI: LCI = .011, UCI = .14), CFI = .923, TLI = .942, and SRMR = .05, which suggest a good model fit, even though the CFI and TLI were a little low. The present model accounted for 49% of the variance in arithmetic fluency. Age, education location, and PMC were significant predictors of arithmetic fluency; participants who were older, educated in Asia, and had higher perceived math competence tended to demonstrate better arithmetic fluency. Of the two quantitative measures, subitizing was the only significant unique predictor of arithmetic fluency (ANS was not). The path from letter naming to arithmetic fluency was not significant, either.

Figure 2.2. Study 1 path model predicting arithmetic fluency.



Note: $^{*}p \leq .05$, $^{**}p \leq .001$, $N = 167$. Standardized path model results shown. Solid lines represent significant paths; dotted lines are specified paths that were not significant. NS = Nonsymbolic. ID = Identification. ANS= Approximate number system. Comp = Comparison. Arithmetic fluency $R^2 = .49$.

Discussion

The subitizing and approximate number systems are considered primitive nonsymbolic quantity systems, and either or both could scaffold the development of the symbolic number system. The goal of Study 1 was to concurrently examine individual differences in subitizing and approximate number skills as predictors of arithmetic fluency in an adult sample. If subitizing is related to mathematics in adulthood, it would lend support to its importance during the acquisition of the symbolic number system.

Study 1 tested three specific hypotheses. The first hypothesis (H1) was that subitizing skill would not be correlated with ANS skill, based on the theoretical claim that these are two distinct nonsymbolic quantitative systems. This hypothesis was supported; the correlation between subitizing and ANS skill was near zero and not significant. The second hypothesis (H2) was that subitizing, but not ANS, would predict arithmetic fluency; this hypothesis was also supported. Although ANS skill was significantly correlated with arithmetic fluency, it failed to account for unique variance in arithmetic fluency with other variables taken into consideration. As far as I know, this is the first study to observe subitizing skill as a significant predictor of mathematics skill in an adult sample, implying that the correlation between subitizing and mathematical outcomes extends across development (i.e., from childhood to adulthood). I believe this is also the first study to include both subitizing and ANS skill as simultaneous predictors of arithmetic fluency. These findings suggest that the subitizing system is relatively more important than the ANS system to mathematical development, supporting the claim that the subitizing system scaffolds the development of the symbolic quantitative system. Note that subitizing was predictive even while controlling for letter naming, suggesting that it is something about the quantitative—rather than speeded—aspect of the subitizing measure that accounted for unique variance in arithmetic fluency.

The third hypothesis tested was that experiential factors would also uniquely predict arithmetic fluency performance (H13 in Table 1.3). This hypothesis was also supported; age, education location, and PMC were significant predictors of arithmetic fluency. In fact, education location was the predictor with the largest standardized beta coefficient—highlighting the importance of including experiential factors when examining mathematical performance.

The present findings refute the ANS-is-Key account, and lend support to the Subitizing-is-Key account. Study 1 did not permit a test of the Subitizing+Counting-is-Key account, as it did not include any nonsymbolic quantitative tasks in the medium range to index the counting system. Furthermore, Study 1 did not include any symbolic quantitative measures. As discussed in Chapter 1, if one or both core nonsymbolic quantitative systems (subitizing and ANS) aid in the development of the symbolic system, they might continue to correlate with mathematical outcomes but fail to account for unique variance when *symbolic* quantitative skills are considered simultaneously. This view is consistent with the recent evidence that performance on symbolic—but not nonsymbolic—quantity comparison tasks uniquely predict mathematical outcomes (De Smedt et al., 2013; see also Schneider et al., 2016).

In order to test the relative importance of symbolic versus nonsymbolic quantitative skills, Study 2 included two kinds of tasks, presented in both symbolic and nonsymbolic formats. The first task-type was rapid identification—that is, the subitizing measure from Study 1, along with an analogous rapid digit naming task. The second task-type was rapid comparison of symbolic and nonsymbolic quantities. The nonsymbolic comparison task included quantities in the small-medium range (1 to 9), rather than the large quantities frequently used in tasks indexing the ANS. Because this task includes the subitizing range, it is my view that it is not a pure ANS task (Dietrich et al., 2015b also state this). Although the ANS might be employed for comparisons of quantities beyond the subitizing range, the counting system might also play a role because the quantities are still relatively small. In sum, the nonsymbolic comparison task in Study 2 was not a pure measure of subitizing, counting, or ANS, and was not labeled as such. It serves as a nonsymbolic task that is analogous to the symbolic comparison task, permitting a test of the relative importance of symbolic versus nonsymbolic quantitative skills to mathematics.

Study 2

Hypotheses

Study 2 tested four of the hypotheses listed in Table 1.3.

- H4: Subitizing will predict symbolic identification (i.e., digit naming) and symbolic comparison.
- H9: Less complex symbolic quantitative skills will predict more complex ones (i.e., digit naming will predict symbolic comparison).
- H10: Nonsymbolic quantitative skills will not predict mathematical outcomes (i.e., arithmetic fluency) beyond symbolic quantitative skills (i.e., symbolic identification and comparison will significantly predict arithmetic fluency, but subitizing and nonsymbolic comparison will not).
- H13: Experiential factors will uniquely predict mathematical outcomes.
- In addition to the overarching hypotheses, subitizing was also expected to predict nonsymbolic comparison because this task included the subitizing range.
- Nonsymbolic and symbolic comparison were also expected to correlate, due to shared task demands.

Method

Participants. Eighty-nine participants completed Study 2. Participants ranged in age from 17 to 54 years and were students attending courses at Carleton University. Median age was 19 years (13 participants did not report their age), 70% were female, and 92% were undergraduates. Approximately 72% had completed their high school education in Canada; 12% in China or Hong Kong, and the remaining 16% completed high school education in other countries. Most participants (70%) reported English as their first language. Other first languages were Chinese (15%), and French (4%). Small percentages (<4%) of other first languages were reported as well (e.g., Italian, Arabic, Farsi; 11% in total)

Procedure. The procedure for Study 2 was the same as that in Study 1. Data collection was supplementary to two laboratory experiments, Gu (2015), and Pusiak, (2015). Fifty-four of the participants in the current study came from Pusiak's Honours thesis which explored differences in the eye-movement patterns associated with addition versus multiplication problems. Thirty-five of the participants in the present study were participants in Gu's Honours thesis, which focused on the addition of dots in grouped versus ungrouped formats (presented on computer). Pusiak (2015) and Gu (2015) administered the secondary tasks in the same order, however, Pusiak administered them prior to the primary experiment (i.e., eye-tracking data collection), whereas Gu administered them afterwards. The secondary task order was as follows: questionnaire, digit naming, subitizing, nonsymbolic comparison, symbolic comparison, arithmetic fluency.

Measures. In the present study the age, education location, subitizing speed, perceived math competence, and arithmetic fluency measures were the same as those in Study 1.

Demographic variables. Age (in years), and education location (1 = educated in Asia, 0 = educated elsewhere) were included as predictors of arithmetic fluency.

Perceived math competence. See Table 2.1 for PMC items, descriptive statistics, and item correlations. Component scores were created via principle components analysis, as was done in Study 1. One component was extracted accounting for 58 percent of the variance. Component loadings are also provided in Table 2.1. The component scores were used in path analyses to predicting arithmetic fluency.

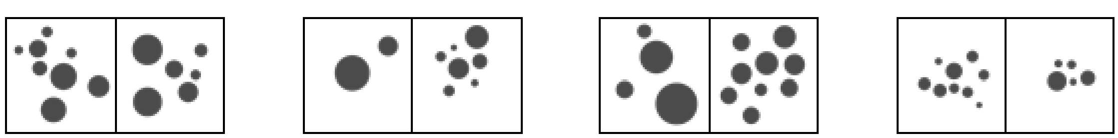
Digit naming (symbolic rapid identification of small quantities). Participants named small digits (i.e., 1, 2, and 3). This task was administered and scored in the same way as the letter naming task in Study 1. The observed parallel-form reliability was .90. Performance is the average items-per-second across the two forms.

Subitizing (nonsymbolic rapid identification of small quantities). This task is the same as that in Study 1. The observed parallel-form reliability was .89 and performance was the average items-per-second across the two forms.

Rapid comparison of quantities. The rapid quantity comparison tasks (symbolic and nonsymbolic) were developed for this study based on similar measures employed by Nosworthy, Bugden, Archibald, Evans, and Ansari (2013). In these tasks, participants saw pairs of numbers or dots and were asked to cross out the greater set in each pair. Nosworthy et al. designed their tasks to be easily administered to large groups of children in a short time frame, without the requirement of any special technology (i.e., a computer or tablet). The present tasks differ from Nosworthy et al.'s in the following ways: 1) In the present versions, quantity comparisons (symbolic and nonsymbolic) included all combinations of 1 to 9 (72 unique pairs considering that each pair is shown two ways, e.g., 5 versus 7 *and* 7 versus 5) although some of these combinations are presented only in the example and practice (see additional information below). Nosworthy et al.'s task included 56 pairs only; they omitted some pairs with redundant ratios while maintaining a range of ratios and balancing right to left presentation across pairs. 2) The present task was administered in two forms (Form A and Form B for both symbolic and nonsymbolic comparison); the difficulty (in terms of ratios) was equated across the two forms. The inclusion of two forms permits parallel-form reliability analyses not possible with Nosworthy et al.'s design. 3) Nosworthy et al.'s tasks start with easy ratios and get progressively more difficult, to minimize discouragement among young children who might struggle with the more difficult ratios. In the present task, the stimuli/ratios were randomly arranged. 4) For the present tasks, time-to-completion and number of errors on each form were recorded; in contrast, participants were given a *set* time limit (i.e., one minute for symbolic and one minute for nonsymbolic magnitude comparison) on Nosworthy et al.'s tasks. Recording time-to-completion with a correction for errors was done to maximize the sensitivity of the measure, particularly among adults who were expected to complete the tasks in less than one minute.

The current tasks included two forms (Form A and Form B) with 30 stimuli on each; the remaining 12 stimuli were included as sample and practice items. The correct response was counter-balanced on each form, such that the correct quantity/digit is on the right for half of the pairs, and on the left for the other half. The stimuli on each form were selected with the ratios of the pairs taken into consideration, as certain ratios are more difficult to discriminate in the nonsymbolic version of the tasks (e.g., discriminating 1 from 9 dots is much easier [i.e., done more quickly and accurately] than discriminating 8 from 9 dots). The difficulty of each form was equated. Furthermore, the dot area was controlled (as was done for Nosworthy et al.'s measure). As some researchers have noted (Gebuis & Reynvoet, 2011), presenting dots of the same size makes it impossible to know whether participants select the greater magnitude based on quantity or total area, as the greater quantity also occupies greater area. Researchers have dealt with this concern by modifying the area of the dots so that participants cannot rely on area alone to be successful in the task. Thus, this task was designed so that area is positively correlated with quantity for half of the stimuli (i.e. the larger quantity also occupies greater area than the smaller quantity), and area is negatively correlated with quantity for the other half (i.e., the smaller quantity actually occupies greater area than the larger quantity). See Figure 2.3 for example stimuli. Once the nonsymbolic stimuli were designed (e.g., ratio-difficulty was balanced across forms; area was controlled for), the dots were simply exchanged for digits representing the same magnitudes to create the symbolic version of the task. See Appendix A for a full list of the stimuli and stimuli characteristics.

Figure 2.3. Examples of nonsymbolic rapid quantity comparison stimuli.



Symbolic comparison of quantity. Participants were shown a practice page with four sample items, one of which was completed. They were told that they would see pairs of numbers (i.e., numerals/digits) and that the objective was to cross out the larger number in each pair, "like this", at which point the experimenter completed the other three sample items. Participants completed eight practice items, and were then told that they would be timed as they completed a whole page of comparisons (i.e., Form A). They were to cross out the numerically larger digit in all of the pairs, starting on the top row and moving from left to right before proceeding to the second row. Once Form A was completed, participants completed Form B. The experimenter recorded the time taken to complete each form in seconds with one decimal, and the form was checked for errors upon completion. Items-per-second scores (with an error correction) were calculated for each form using the same formula employed for the subitizing and digit naming measures. The observed parallel-form reliability was .91. Performance was the average of the items-per-second scores on the two forms.

Nonsymbolic comparison of quantity. This task was administered and scored in the same way as the symbolic comparison task—the difference being that participants saw pairs of dots rather than digits. They were asked to cross out the set in each pair with more dots. Participants completed Forms A and B and the experimenter recorded the time taken to complete each form. Each form was checked for errors upon completion using a correctly completed version of the task printed on a transparency. Items-per-second scores (with an error correction) were calculated for each form using the same formula as above. The observed parallel-form reliability for the rapid magnitude comparison task-nonsymbolic was .89. Performance was the average of the two item-per-second scores.

Arithmetic fluency. The Calculation Fluency Test described in Study 1 was the dependent measure in Study 2. Performance was total correct across the three subtests.

Results

Descriptive statistics for Study 2 are shown in Table 2.5. The table displays raw scores for the rapid identification and rapid comparison tasks. The summary scores for these measures are also shown (e.g., the mean items-per-second scores). Each of these measures had two forms; the minimum and maximum values are collapsed across the two forms. Comparisons across tasks between Study 1 and Study 2 are shown in Appendix B. In general, performance was very similar on the common tasks between the two studies.

Correlations. Correlations among variables included in the Study 2 path analyses appear in Table 2.6. Education location was fairly highly correlated with arithmetic fluency, indicating that Asian-educated participants performed significantly better on this measure relative to those educated elsewhere. Age was positively correlated with education location, and marginally correlated with arithmetic fluency. Perceived math competence (PMC) was moderately correlated with arithmetic fluency, and with nonsymbolic comparison. All four quantitative tasks—subitizing, digit naming, symbolic and nonsymbolic comparison—were moderately correlated with arithmetic fluency, as well as with one another. The analogous tasks (i.e., symbolic and nonsymbolic versions of the same tasks) were highly correlated, as was expected given the similar task demands and administration procedures.

Table 2.5.

Study 2 Descriptive Statistics (PMC items excluded)

	Raw Scores								Summary Scores: Items-per-second				Arithmetic Fluency (CFT)		
	Digit Naming		Subitizing		Symbolic Comparison		NS Comparison		Digit Naming	Subitizing	Symbolic Comparison	NS Comparison	Add.	Sub.	Mult.
	Time	Err.	Time	Err.	Time	Err.	Time	Err.							
Min	4.2	0	5.6	0	10.7	0	12.7	0	2.0	1.8	.7	.5	4	0	0
Max	12.6	2	14.3	2	47.7	3	54.1	6	5.6	4.2	2.7	2.4	0	25	24
M	7.5	.1	9.2	.1	20.5	.2	27.1	.9	3.3	2.7	1.5	1.2	15.0	10.7	6.4
SD	1.6	.4	1.7	.3	5.0	.5	7.0	.9	.71	.49	.34	.32	6.3	5.6	5.8

Note: $N = 89$. NS = Nonsymbolic. Err. = number of errors. CFT = Calculation fluency test; scores are total correct. Add. = Addition subtest. Sub. = Subtraction subtest. Mult. = Multiplication subtest. Time is measured in seconds. See Table 2.1 for PMC descriptive statistics, item correlations, and component loadings.

Table 2.6.

Correlations Among Variables Included In Study 2 Path Analyses

	1.	2.	3.	4.	5.	6.	7.
1. Age (years)	--						
2. Education Location	**.29***	--					
3. PMC	-.07	.01	--				
4. Digit Naming	.01	.14	.18	--			
5. Subitizing	.07	.15	.15	**.76****	--		
6. Symbolic Comparison	.09	**.21***	.11	**.49****	**.57****	--	
7. NS Comparison	.01	.13	**.22***	**.46****	**.51****	**.71****	--
8. Arithmetic fluency (CFT)	**.21**†	**.53****	**.34****	**.48****	**.57****	**.57****	**.43****

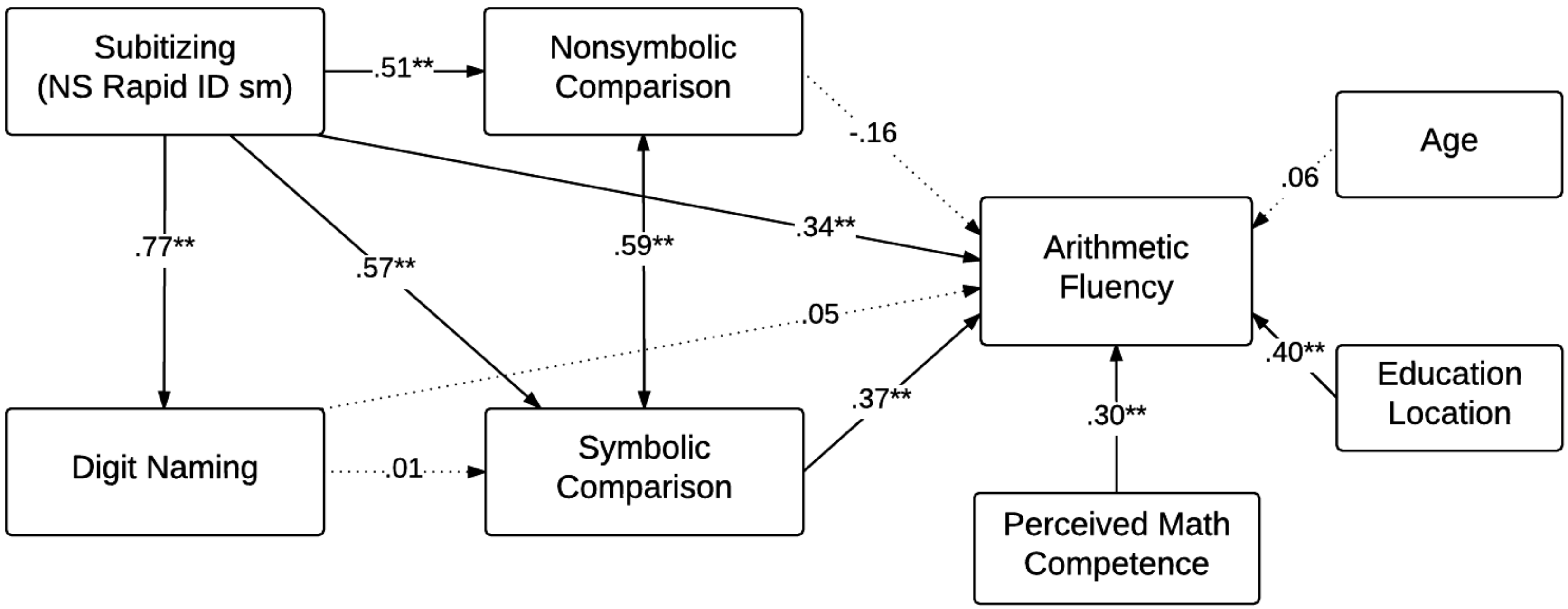
Note: $^{\dagger}p = .07$; $^{*}p \leq .05$, $^{**}p \leq .01$. $N = 76$ for correlations with Age, and 89 for all other correlations. Education location is coded 0 = Educated elsewhere, 1 = Asian-educated. PMC = Perceived math competence. NS = Nonsymbolic. CFT = Calculation fluency test. Table 2.1 contains PMC items, descriptive statistics, and PCA results.

Path model. Path analysis was conducted in Study 2 to test the stated hypotheses. Unlike the path model in Study 1, this current model included both exogenous and endogenous[4] predictors of arithmetic fluency. The hypothesized paths are shown in Figure 2.4, along with the standardized solution, as produced in MPlus. All model fit indices suggested this path model was a very good fit to the data, $\chi^2(10) = 9.85$, $p = .45$, RMSEA= .000 (LCI = .000, UCI = .012), CFI = 1.00, TLI = 1.00, SRMR = .042.

Two of the four hypotheses for Study 2 were supported. Subitizing was a significant predictor of symbolic identification and comparison (H4). Perceived math competence and education location uniquely predicted arithmetic fluency (H13), although age did not. As expected, subitizing predicted nonsymbolic comparison, presumably because the comparison task included the subitizing range, and nonsymbolic comparison was significantly correlated with symbolic comparison. However, two hypotheses were not supported. The hypothesis that less-complex symbolic skills would predict more-complex ones (H9) was not supported as digit naming did not predict symbolic comparison. Nonsymbolic quantitative skills were not expected to predict mathematical outcomes once symbolic skills were taken into consideration (H10). This hypothesis was not supported either, as subitizing, but not digit naming, uniquely predicted arithmetic fluency. Both subitizing and symbolic comparison predicted arithmetic fluency, supporting the view that the core quantitative skill of exact quantity identification continues to be related to mathematical skill in adults. Presumably both subitizing and symbolic number skills form part of the quantitative pathway in the larger model of mathematical skill (LeFevre et al., 2010).

[4] Endogenous variables are predicted by variables within the model, whereas exogenous variables are not (i.e., are caused by external variables). The variance accounted for in endogenous variables is known.

Figure 2.4. Study 2 path model: Quantitative and experiential predictors of arithmetic fluency.



Note. *$p \leq .05$. **$p \leq .01$, $N = 89$. Standardized solution shown. Dotted arrows represent nonsignificant paths that were hypothesized in the model specification. NS = Nonsymbolic. Arithmetic fluency $R^2 = .68$.

The estimated R^2s (variance accounted for) for in endogenous variables (i.e., variables receiving arrows in Figure 2.4) were as follows: Arithmetic fluency, $R^2 = .68$; digit naming, $R^2 = .59$; symbolic comparison, $R^2 = .33$; and nonsymbolic comparison, $R^2 = .26$. All R^2s were highly significant (ps = < .001 to .003). The standardized indirect effects of digit naming and subitizing on arithmetic fluency were tested. There was no indirect effect (effect = .00) of digit naming (symbolic identification). The indirect effect of subitizing on arithmetic fluency was .17 (primarily through symbolic comparison), which was marginally significant ($p = .065$); when added to the direct effect, the total effect of subitizing was .51, making it the greatest overall contributor to arithmetic fluency.

Discussion

Study 2 included analogous symbolic and nonsymbolic identification and comparison tasks, and the hypothesized relations among these quantitative skills were tested. Quantitative skills were also examined as unique predictors of arithmetic fluency. Both the Subitizing-is-Key and Subitizing+Counting-is-Key accounts claim that subitizing scaffolds verbal counting and symbolic number system knowledge. Consistent with these views, subitizing was observed to predict symbolic identification and symbolic comparison.

Subitizing was found to be a significant predictor of the nonsymbolic comparison measure. As mentioned above, the nonsymbolic comparison task in the current study was problematic because the contributions of the subitizing, counting, and approximate number systems cannot be disentangled. Study 3 addressed these problems in two ways: First, the small, medium, and large quantity ranges were defined with no overlap. It was assumed that nonsymbolic tasks in these ranges isolate the subitizing, counting, and approximate number systems respectively. Isolating the quantitative systems permits formal tests of the hypothesized relations among them, as well as tests of their relative contributions to adults' mathematical outcomes. Second, Study 3 included identification *and* comparison tasks—as well as ordering tasks—across all three ranges. If nonsymbolic identification, comparison, and ordering tasks within a range correlate and load onto the same factor, it will suggest that these tasks tap the same underlying construct (i.e., counting skill in the medium range). In contrast, if tasks do not group together as expected, it will suggest that different types of tasks require different forms of processing (e.g., nonsymbolic comparison medium [5 to 9] employs different underlying numerical representation[s] than nonsymbolic identification in the same range).

Symbolic comparison was a unique predictor of arithmetic skill as hypothesized. Contrary to expectations, subitizing not only correlated with arithmetic fluency (as it did in Study 1), but was a unique predictor of arithmetic fluency, whereas symbolic identification (i.e., digit naming) was not. (Subitizing also predicted symbolic comparison whereas symbolic identification did not). In fact, when direct and indirect effects were considered (total standardized effect = .51), subitizing was the strongest overall predictor of arithmetic fluency.

In sum, the findings in the present chapter provide the first evidence that subitizing is predictive of arithmetic fluency among adults. In Study 1, subitizing skill predicted arithmetic fluency, while ANS skill did not, suggesting the subitizing makes a greater relative contribution to mathematical skill development. In Study 2, I expected that relations between subitizing and arithmetic fluency would be mediated by symbolic quantitative skills, however, subitizing predicted symbolic quantitative skills, and also accounted for unique variance in arithmetic fluency (along with symbolic quantity comparison performance).

Study 3 included symbolic and nonsymbolic quantity identification, comparison, and ordering in three distinct ranges (small, medium, and large). This comprehensive set of tasks was designed to create pure indexes of the subitizing, counting, and approximate number systems. It was then possible to examine hypothesized relations among these systems (i.e., subitizing was expected to predict counting) and symbolic quantitative skills (i.e., subitizing and counting were expected to predict symbolic identification and comparison). Symbolic and nonsymbolic quantitative skills and experiential factors were simultaneously examined as unique predictors of multiple mathematical outcomes, while also taking domain-general cognitive abilities into consideration.

CHAPTER 3: STUDY 3

The purpose of the present study was to systematically examine theoretically relevant quantitative processes in relation to each other and as predictors of mathematical outcomes in adults. The hypotheses tested in Study 3 are summarized in Table 3.1.

Table 3.1. *Summary of Study 3 Hypotheses*

Category	Hypotheses
NS sm (1 to 3): **Subitizing**	1. Subitizing skill will not relate to ANS skill 3. Subitizing skill will predict counting skill 4. Subitizing skill will predict symbolic identification and comparison
NS lg (10 to 100): **ANS**	5. ANS skill will not uniquely predict *exact* mathematical calculation 6. ANS skill will predict number line performance
NS med (5 to 9): **Counting**	7. Counting will predict symbolic identification and comparison
Tasks-type and format	8. Nonsymbolic identification, comparison, and ordering tasks from the same range (small, medium, and large) will tap the same underlying constructs (subitizing, counting, and ANS systems) 9. Less complex symbolic quantitative skills will predict more complex ones (i.e., identification will predict comparison, comparison will predict ordering) 10. Nonsymbolic quantitative skills will not predict exact arithmetic-based outcomes beyond symbolic quantitative skills (cf. H6) 11. Symbolic ordering will be the best quantitative predictor of mathematical outcomes (except for maybe NS number line)
Domain-general skills	12. Domain-general cognitive skills will predict domain-specific quantitative skills and have indirect effects on mathematical outcomes
Experience	13. Experiential factors will uniquely predict mathematical outcomes

Note: Hypothesis 2 only applied to Study 1.
NS = Nonsymbolic. sm = small, med = medium, lg = large.

The quantitative tasks selected for inclusion in this study varied on three dimensions: format (symbolic [digits] or nonsymbolic [dots]); range of the quantities (small [1 to 3], medium [5 to 9], and large [10 to 100]); and task (identification, comparison, and ordering). Table 3.2 provides a summary of quantitative measures in Study 3, as well as the constructs they were intended to assess. In the past, researchers tended to measure subitizing and counting via nonsymbolic identification in the small and medium ranges respectively. In contrast, ANS skill has typically been assessed via nonsymbolic comparison tasks in the large range (cf. Lyons & Beilock, 2011; Nosworthy et al., 2013). Thus, the range (sizes of the quantities) and task-type (identification and comparison) have been partially or wholly confounded across studies. In contrast, the present research was designed to systematically include identification and comparison tasks within the small, medium, and large ranges. Order judgement tasks were also included as they have been employed in recent research. Nonsymbolic measures within the small, medium, and large ranges were expected to tap the subitizing, counting, and ANS systems, respectively, whereas symbolic quantitative measures were expected to tap one underlying numerical representation. The inclusion of the same tasks in both symbolic and nonsymbolic formats was motivated by research findings that show that ANS skills do not predict mathematical outcomes once symbolic skills are taken into consideration.

In the present study quantitative measures were examined in relation to each other, in relation to domain-general skills, and as predictors of different mathematical outcomes, including arithmetic fluency, calculation knowledge, and number line estimation.

Table 3.2.

Summary of Domain-Specific Quantitative Measures (detailed)

Format: Symbolic					Format: Nonsymbolic				
		Task					Task		
Construct	Range	ID	Comp	Order	Construct	Range	ID	Comp	Order
Symbolic Quantitative Skills	Small (1-3)	✓	✓		Subitizing	Small (1-3)	✓	✓	
	Medium (5-9)	✓	✓		Counting	Medium (5-9)	✓	✓	
	Small-Medium (1 – 9)			✓	Subitizing/ Counting	Small-Medium (1 – 9)			✓
	Large (10 – 100)	✓	✓	✓	ANS	Large (10 – 100)	✓	✓	✓

Note: ID = Identification; Name the digit/quantity. Comp = Magnitude Comparison; Which digit/quantity is greater? Order = Ordering, Are these digits/quantities in ascending order? ANS = Approximate number system.

Method

Procedure

Participants completed the study in exchange for 2% credit in a psychology course. To recruit participants, the study description and available timeslots were posted to Carleton University's online research participation system (SONA), and participants signed up to participate. One of three female experimenters administered the testing session in a university-based research laboratory. Participants completed paper-and-pencil, iPad, and computerized tasks. Computerized stimuli were presented with the software program, Superlab. See Table 3.3 for a summary of task order and forms of administration. Detailed task descriptions are provided below.

Participants

Participants were 153 undergraduate students (50.3% females) from Carleton University. Ages ranged from 17 to 56 years; the median age was 21 years (SD = 6.1). Sixty-eight percent of participants reported English as their first language. Other first languages reported were Chinese (8.5%), and Arabic (9.2%), as well low frequencies of other languages.

Measures

The measures included in the present study are described here. At the end of each description, the indicator of performance is stated (e.g., P, absolute error, RT). The scoring details and rationale are provided under the Data Processing heading in the Results section. Note that P is an error-corrected RT; the original RTs are increased monotonically with the rate of error. Task reliabilities are also provided in the descriptive statistics tables in the Results section.

Table 3.3.

Study 3 Task Order and Method of Administration

Tasks	Task Administration
1. Symbolic Rapid Identification Small (1- 3) 2. Nonsymbolic Rapid Identification Small (1- 3)	Paper-and-pencil
3. Symbolic Identification (Small to Large) 4. Symbolic Comparison (Small to Large) 5. Symbolic Ordering (Small to Large) 6. Nonsymbolic Identification (Small to Large) 7. Nonsymbolic Comparison (Small to Large) 8. Nonsymbolic Ordering (Small to Large) 9. Symbolic to Nonsymbolic Mapping 10. Speeded Addition	On computer, using Superlab software
11. Calculation Fluency Test 12. Brief Math Assessment-III	Paper-and-pencil
13. Symbolic Number Line 14. Nonsymbolic Number Line 15. Visual-Spatial Span	iPad applications
16. Choice Response Time 17. Go/No-Go 18. Black White Stroop	On computer, using Superlab software
19. Backward Digit Span 20. WASI Vocabulary (Verbal IQ) 21. WASI Matrix Reasoning (Nonverbal IQ)	Paper-and-pencil
22. Math Background and Information Questionnaire (MBIQ)	On computer, using web survey tool

Quantitative skills. Analogous quantitative skills were administered in symbolic and nonsymbolic formats. In the following section the shared administration details are provided first, followed by details that varied by format. A paper-and-pencil administration was employed for the rapid identification of small quantities tasks; the identification, comparison, and ordering tasks were administered on a computer using the stimulus presentation software, Superlab.

Rapid identification small. Both the symbolic and nonsymbolic rapid identification tasks began with instructions and six practice items. For each form, participants identified (named) all of the stimuli as quickly as possible, moving from left to right, and then to the next line, etc. The experimenter recorded the time taken to name all stimuli on a form, as well as the number of errors. Participants completed two forms with 24 stimuli each. Performance is the mean items-per-second score across forms.

Symbolic rapid identification small. Participants named the digits "1", "2", and "3" as quickly as possible; stimuli were presented randomly. This task was the same as the digit naming task employed in Study 2.

Nonsymbolic rapid identification small. Participants identified (named) 1 to 3 dots enclosed in boxes as quickly as possible; stimuli were analogous to those in the symbolic rapid identification task. Dots were not presented in a canonical format (i.e., as on dice). This was the same subitizing measured employed in Studies 1 and 2.

Identification small, medium, and large. Participants were told that they would see stimuli (numerals or dots) and they were to name them as quickly as possible. After instructions, participants completed a practice block, followed by a small-medium block (1 to 9), and then a large block (10 to 100). In the small-medium block each stimulus (1 to 9) was presented four times for a total of 36 trials. The set of 53 large stimuli were: 12, 13, 14, 15, 18, 19, 21, 23, 25, 27, 28, 29, 32, 34, 35, 36, 39, 42, 43, 45, 47, 48, 49, 51, 52, 54, 56, 57, 59, 61, 63, 64, 65, 68, 69, 71, 72, 73, 75, 76, 78, 81, 82, 84, 85, 87, 89, 91, 93, 95, 96, 97, and 98. Stimuli in each block were presented in random order. Each stimulus remained on the screen for a maximum of 2000ms, or until the participant responded. Response times were recorded via voice key; responses were entered into a text box by the experimenter. As soon as the experimenter hit Enter, the next trial began.

Symbolic identification. Participants completed four practice trials. Performance is median RT by participant in the small (1 to 3), medium (5 to 9) and large (10 – 100) ranges.

Nonsymbolic identification. Participants were told that they would see 1 to 100 dots and they were to say how many there were as quickly as possible. The instructions indicated that they might have to guess the number of dots at times. There were 12 practice trials (3, 7, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100). Practice stimuli were selected to span the whole range of quantities and were presented in random order. All dots were red in colour; a red-white perceptual mask appeared on the screen for 50ms after the dot presentation, followed by the text box (into which the experimenter entered the response). For small nonsymbolic identification, performance was median RT for trials with 1 to 3 dots. *P* was the performance measure for nonsymbolic identification in the medium range (5 to 9 dots). Nonsymbolic identification large was mean percent absolute error (calculated with the aggregate function in SPSS).

Comparison small, medium, and large. The comparison tasks began with instructions and practice trials to familiarize participants with the procedure. Each trial began with a fixation cue. A pound symbol (i.e., #) flashed twice in the middle of the computer screen (# for 500ms, 200ms pause, # for 500ms), followed by a two numerals or sets of dots presented side-by-side. Stimuli remained on the screen for 2000ms or until the participant responded. Participants were instructed to indicate the greater quantity. The one on the right was selected by pressing the "J" keyboard button with the right index finger; the one on the left by pressing the "F" keyboard button with the left index finger. After a response, or when the trial timed out, a perceptual mask appeared for 50ms before the next trial began. The first test block (87 trials) consisted of small-medium quantity comparisons (1 to 9). The second block included comparisons in the large range (10 to 100; 108 trials). Prior to beginning each block, participants were informed of the range of quantities included. The side with the greater quantity was counterbalanced such that it was on the right for half of trials, and on the left for the other half. Each stimuli pairing was presented twice; once with the greater quantity on the right, and once with the greater quantity on the left (e.g., 3 5 and 5 3 were presented). The full set of small-medium comparison stimuli is shown in Appendix C.

Large comparison stimuli were designed to include the same ratios as the small-medium stimuli, along with some more difficult ratios (an important consideration for nonsymbolic comparison). More variance was expected in nonsymbolic comparison relative to symbolic. Because subitizing was an important construct in the present study, the nonsymbolic comparison task was designed to include more comparisons in the small range than in the symbolic version, in order to promote a stable (subitizing) RT. There were also a few idiosyncratic differences between the symbolic and nonsymbolic comparison tasks due to programming errors when setting up the stimuli presentation. A full list of large comparison stimuli is provided in Appendix D.

Symbolic comparison (SC). The symbolic comparison task included four practice trials. The numerals were displayed in black, 96-point, Arial font on a white background. A black-white perceptual mask was shown after each pair of digits.

Nonsymbolic comparison (NSC). Nonsymbolic comparison included six practice trials. For every trial, one set of dots was purple and the other green to promote distinction between the two sets. A green-purple perceptual mask was shown after each pair of dots to minimize the effect of the retinal after-image created by the stimuli.

Nonsymbolic comparison stimuli were designed to vary on perceptual cues typically associated with numerosity, thus minimizing the likelihood that participants could rely on such cues to be successful. Dots in each set appeared in various sizes, as a consistent dot size might provide such a cue. Consistent with other recommendations to control for visual properties of the stimuli (Gebuis & Reynvoet, 2011), half of the trials consisted of congruent trials (area correlated or AC) and half consisted of incongruent trials (perimeter controlled or PC). Area was correlated with quantity in congruent trials; the greater number of dots took up more overall area than the lesser set. In incongruent trials, the perimeter of the two sets of dots was controlled and the lesser quantity occupied more overall surface area than the greater set. Thus, respondents had to inhibit area information and focus on numerosity to be correct on incongruent trials. The stimuli list in Appendix D indicates which trials were congruent or incongruent.

Ordering small-medium and large. Each ordering trial consisted of a fixation cue (500ms), a pause (500ms), and three quantities (up to 2000ms), followed by a perceptual mask (50ms). The stimuli remained on the computer screen for 2000ms or until the participant responded. The task began with instructions. Participants were informed that in each trial they would see three numbers/sets of dots and that they were to indicate whether they were in ascending order or not (from smallest to largest/least to greatest). They were to push the keyboard button "K" with their right index finger if the numbers were *in order* and the keyboard button "D" with their left index finger if the numbers were *not in order*. They were encouraged to be quick and accurate. The task began with four practice trials. After the practice, participants completed the first block of test trials; they were informed that this block contained quantities 1 to 9 only (the small-medium range). After the small-medium block, participants were shown another instruction screen which informed them that the next block contained large quantities only (10 to 100); task instructions were also reiterated on this screen. Each test block consisted of 84 trials which were presented randomly. Half of the trials were in order and half were not in order. None of the trials were in descending order. For all ordered trials, there was an equivalent unordered trial with the same three quantities (e.g., 1 2 3 appeared, as did 2 3 1). The large ordering stimuli were designed to be analogous to the small-medium stimuli in that ratios were roughly equivalent. A list of ordering stimuli can be found in Appendix E.

Symbolic ordering (SO). In the symbolic ordering task, the fixation cue was a box that appeared on the screen around the area where the numerals were about to appear. The perceptual mask shown after the ordering stimuli was black and white, as numerals were black on a white background.

Nonsymbolic ordering (NSO). In this task participants saw three sets of dots, and each set was presented in its own respective area on the computer screen (i.e., one third of the screen; dot sets were not intermixed). The fixation cue consisted of three asterisks (*), each one presented in the middle of each dot area. To aid discrimination, the three sets of dots in each trial appeared in three different colours. The colours were red, purple, and green. The colours changed position randomly from trial to trial (i.e., left, middle, right). A red-purple-green perceptual mask (50ms) was used in this ordering task. To prevent participants from relying on area alone to make their decisions, the set of dots occupying the greatest area was counterbalanced across locations. That is, the largest quantity occupied the greatest area, the medium area, and the smallest area for one third of the trials (as did the medium and smallest quantities).

Domain-general cognitive skills: Executive function. Choice response time, visual spatial span, backward digit span, black white Stroop, and go/no go tasks were administered. These tasks were used as indicators for an executive function factor, representing domain general skills.

Choice RT. A choice response time task was the measure of attention and processing speed. Xs and Os were presented with a computer. Participants pressed the X on the keyboard with their left index finger when they saw X, and the O on the keyboard with their right index finger when O appeared. Trials began with a fixation point (500 ms), followed by the letter until the participant responded or for a maximum of 2000 ms. The next trial began immediately after a response. There were 50 trials in total (25 Xs and 25 Ox). Stimuli were presented in a set order designed to seem random (i.e., to prevent long strings of Xs or Os as could occur if stimuli were randomized).

Black White Stroop. Vendetti, Kamawar, Podjarny and Astle (2015) developed this task based on the one designed by Gerstadt, Hong, and Diamond (1994). Participants were required to suppress a prepotent (i.e., habitual) response and thus this task is assumed to index inhibitory control. Stimuli were black and white boxes on a grey background, presented on a computer screen. Participants were instructed to say "black" when they saw a white box and "white" when they saw a black box. Response times were detected by a voice key, and the experimenter entered actual responses with keyboard buttons. The score on this measure is median RT on correct trials.

Go/No Go task. A Go/No-Go task was used as a second measure of inhibitory control. This was a continuous task in which participants had to keep track of the letter they had previously seen in order to respond correctly. Each trial began with a fixation cue comprised of three dots (i.e., an ellipsis) presented on a computer screen for 200 ms, followed by a letter (X or Y) for a maximum of 1500 ms, or until a response from the participant. The cue and letter were presented in black, 72-point Geneva font. The cue appeared between each presentation of a letter. If the current letter was *different* from the previous letter (e.g., X followed by Y), participants were instructed to press the space bar. If the current letter was the *same* as the previous letter (e.g., X followed by X), they were to do nothing. Letters were not paired, but appeared as a consecutive string; participants had to constantly keep the last letter in mind. Because participants were much more likely to see two different letters in succession than two same letters, the prepotent response was to press the space bar. Thus, participants had to inhibit their prepotent response when two of the same letters appeared sequentially.

Before participants began the practice trials, the experimenter pointed out that the letters were successive—not paired—and thus the previous letter had to be remembered at all times. Participants were told to respond quickly as trials might time-out otherwise. The task began with six practice trials. All trials (including practice) included feedback. That is, if the participant pressed the space bar when s/he was not supposed to (e.g., when X followed X), s/he was informed that this was incorrect. The error message also included a reminder of the last two letters, so that one could continue successfully once the task resumed (i.e., respond correctly when the next letter came up). The test consisted of 124 letter presentations. Of these, 104 (84%) were Go trials (i.e., pressing the space bar was correct) and 20 (16%) were No-Go trials (i.e., doing nothing was correct). Consistent with signal detection theory, and with the literature (e.g., Clark, et al., 2013), the variable derived from this task is response accuracy or d' ($d' = z$[correct Go trials] – z[false alarms on No-Go trials]). In this scoring "false alarms" (i.e., pressing the space bar on No Go trials) are subtracted from "hits" (correct Go trials).

Visual Spatial span. A visual-spatial span task modeled after the Corsi blocks task (Berch, Krikorian, & Huha, 1998) served a measure of visual spatial attention (Rasmussen & Bisanz, 2005). This task was administered with an iPad. Participants saw nine green circles, and watched as they lit up sequentially in random patterns. They then touched the circles in the same order as they lit up during the trial. This task requires the resources of the visual-spatial sketchpad, since participants much keep track of locations and the sequence with which they light up (LeFevre et al., 2010); that is, it is a measure of spatial working memory (LeFevre, DeStefano, Colemand & Shanahan, 2005).

The first trial (sequence length of 2) was a practice trial. Test trials consisted of three sequences of sequence lengths from two to nine. The testing application was programmed to stop once the participant had completed all 24 spans or when a participant made three errors within the same span (e.g., if a participant made errors on all three trials with a span of seven, they would not get to attempt spans of eight). Performance was scored as the total correct number of sequences; the maximum score was 24.

Backward Digit span. Participants heard a string of numbers and were asked to verbally repeat the numbers (single digits) in reverse order. The task began with sequences of two numbers. Sequence length increased by one number at a time, up to a maximum of nine; there were two trials of each sequence length. The task ceased when participants were incorrect on both sequences of the same span. Performance was scored as the total number of sequences reproduced correctly and the maximum score was 16. This task requires resources from the phonological loop, however, it is also assumed to also require the resources of the central executive, since the order of the digits must be manipulated during retrieval (Baddeley, 2001).

Experiential factors. Participants self-reported their demographic information (e.g., gender, language[s] spoken) and completed the Math Background and Information (MBIQ) survey. This survey included questions to assess perceived math competence, which was included as an experiential factor, along with age[5] and education location.

---

[5] Gender was considered as a potential predictor of mathematical outcomes. It significantly correlated with Speeded Addition (-.24), BMA (-.22) and Symbolic Number Line (-.20); males outperformed females on these tasks. However, gender did not uniquely predict math outcomes in SEM analyses. For parsimony and to aid model fit (i.e., decrease the number of free parameters), gender was not included in further analyses.

Age. Age in years was included as a predictor of mathematical outcomes. It was considered an experiential variable because the correlation between age and mathematical performance is believed to be a cohort effect, rather than an improvement in math skills with age. That is, older adults are presumed to have had different educational experiences, including more arithmetic practice relative to younger participants, and these experiences are responsible for the correlation between age and mathematical outcomes. Consistent with this view, LeFevre et al. (2014) found that arithmetic fluency among young Canadian adults declined steadily between the years 1993 and 2005.

Education location. As in Studies 1 and 2, a dichotomous Education Location variable was created in which *Asian-educated* = 2, and *Educated elsewhere* = 1, thus positive correlations/regression coefficients reflect higher performance among Asian-educated participants relative to others. Asian nations are top-ranked on international math assessments and there is evidence that there is greater emphasis placed on math proficiency within these cultures. Thus, this variable was included to capture individual differences in mathematical outcomes associated with differences in learning experiences in Asian countries versus elsewhere.

Perceived math competence. Participants self-reported (on a 7-point-scale) their perceived math competence (PMC; items are the same as in Studies 1 and 2). Table 3.4 includes item wording, response scale anchors, descriptive statistics, and observed internal reliability. In the present study, PMC items were entered into a confirmatory factor analysis to create a PMC factor (details below).

Table 3.4.

Perceived Math Competence: Descriptive Statistics Study 3

Items	*N*	Min	Max	*M*	*SD*
1. Please rate your level of basic mathematical skill (e.g., skill at arithmetic)	153	1	7	5.0	1.38
2. Please rate your level of mathematical skill in more complex areas of mathematics (e.g., calculus, algebra)	153	1	7	4.1	1.64
3. How often do you avoid situations involving mathematics?	152	1	7	4.6	1.75
4. How often do you find that situations involving mathematics make you nervous?	151	1	7	4.7	1.78
5. How difficult was mathematics for you in high school?	153	1	7	4.8	1.71
Reliability (Cronbach's α) = .83					

Note: For all items, participants responded with the numbers from 1 to 7. For Items 1 and 2, response options were 1 = *Very Low*, 4 = *Moderate*, and 7 = *Very High*. Response options for Items 3 and 4 were 1 = *Almost Always*, 4 = *Sometimes*, and 7 = *Almost Never*. For Item 5, response options were 1 = *Extremely Difficult*, 4 = *Moderately Difficult*, and 7 = *Not at All Difficult*. Responses in between anchors were not labeled.

Dependent variables

Arithmetic fluency. The calculation fluency test (CFT) was used to measure arithmetic fluency (CFT; Sowinski et al., 2014). The CFT is a timed multi-digit arithmetic task. There are three separate forms. Each form contains one operation (addition, subtraction, or multiplication) and participants get one minute per form to do as many problems as they can. Performance is total correct on each form. The three subtests were entered as indicators in a CFA to create an arithmetic fluency factor.

Speeded addition. A computerized addition task was administered via Superlab. The task began with instructions. Participants were informed that they would answer addition questions and were encouraged to respond quickly. Participants input responses with a number keypad; a text box on the screen displayed their answer. They pressed the Return key to enter each response and move on to the next problem. The addition task consisted of three practice trials, followed by a block of small addition (16 trials; sums less than 10), then a block of larger addition trials (16 trials; operands 2 to 8; sums between 10 and 16). Accuracy and RTs were recorded. Performance is P (collapsed across blocks).

Calculation knowledge (the Brief Math Assessment). Participants completed the 10-item Brief Math Assessment (BMA-3; Steiner & Ashcraft, 2012), which was developed from a standardized math test called the Wide Range Achievement Test Third Edition (WRAT3; Jastak Associates, 1993). It covers arithmetic with whole numbers and fractions, as well as algebraic knowledge and simplification of terms; it begins with simple items and gets progressively more difficult. Participants were given extra paper to work out problems, but were not permitted technological aids (i.e., calculators). Unlike the calculation fluency test, this measure was not speeded. Eight minutes were allotted to this task for scheduling purposes. Most participants finished within this time; additional time was provided if desired. Performance was total correct; maximum score is 11.

Symbolic number line task (1 to 1000). This number-to-position number line task was originally developed by Siegler and Opfer (2003). In this study, it was administered using an iPad App called EstimationLine that was purchased through the iTunes store. This application can be configured to meet various research needs (e.g., the line endpoints can be changed; a midpoint can be added; stimuli can be changed). The task began with written and verbal instructions. Participants were told that they would view a line with zero on the left end and 1000 on the right end. They were informed that they would see target numbers that they were to place on the line by touching the line to select their chosen location. The tasks began with three calibration trials in which a hatch line was shown on the 0 to 1000 number line, and participants were asked to touch the hatch line. When they touched it, another hatch line of a different colour appeared. The calibration trials allowed the participant to calibrate their proficiency at touching the line (i.e., touching the line *where* they wanted to touch it).

The calibration trials were followed by the test trials. Participants placed target numbers on the line by touching the line in the selected position; when they touched the line, a hatch mark appeared. The symbolic number line task had 24 trials. The target numbers were: 6, 18, 59, 124, 165, 211, 239, 344, 383, 420, 458, 500, 542, 580, 1617, 656, 761, 789, 835, 876, 903, 982, 991, and 994. Target numbers were presented in a randomly generated order. Performance was calculated in three ways: Mean percent absolute error, linear slope, and R^2 of the line. Percent absolute error was calculated for each trial (PAE = |target - response|/1000). Mean percent absolute error per participant was calculated using the aggregate function in SPSS. Participants' responses were plotted against the actual numerical values to fit the linear function, thus producing a slope the variance accounted for (R^2) for each participant. To summarize, three indicators of performance were derived from the symbolic number line task; these three measures were employed as indicators to create a symbolic number line factor.

Nonsymbolic number line task (1 to 100). This task was based on a similar task used by Sasanguie, De Smedt, Defever, and Reynvoet (2012; see also Kolkman, et al., 2013). The nonsymbolic number line task directly followed the symbolic version, and was administered with the iPad in a very similar way. It also began with instructions and three calibration trials. Participants were informed they would see a number line again, but this time there would be zero on the left end and 100 dots on the right end.[6] They were told they would sets of dots (i.e., target quantities) and were to place them on the number line by touching it. When they touched the line, a hatch mark appeared. The nonsymbolic number line task had 26 trials. The target quantities were: 2, 3, 5, 8, 14, 17, 21, 26, 32, 39, 45, 47, 50, 53, 55, 61, 68, 74, 79, 83, 86, 92, 95, 97, 98, and 100. Nonsymbolic number line performance was calculated in the same three ways as symbolic performance: Mean percent absolute error, linear slope, and R^2 of the linear fit. These performance measures were indicators of a nonsymbolic number line factor (details to follow).

[6] This task may seem odd, but in general participants grasped it quickly—probably because they had just completed the symbolic number line task and could see how this task was an analogous version of that task, but with nonsymbolic stimuli.

Omitted measures. Participants completed measures of verbal and nonverbal intelligence (the Vocabulary and Matrix Reasoning subscales of the Weschler Abbreviated Scale of Intelligence [WASI-II]; Wechsler, 2011). Participants produced definitions of target words in the Vocabulary subtest. The Matrix Reasoning subscale required participants to select the picture missing from a visual pattern or sequence. The verbal and nonverbal intelligence measures were correlated with executive function, and some of the quantitative measures. There were few correlations between the intelligence measures and the mathematical outcomes, and intelligence measures did not uniquely predict math outcomes. It was decided that the executive function factor sufficiently captured domain-general cognitive skills for the purposes of the present research, and thus the verbal and nonverbal intelligence measures were deemed superfluous and excluded from further analyses.

Participants also completed a task intended to tap individual differences in symbolic-to-nonsymbolic quantity mappings in the large range. Participants were shown two-digit numerals, and had to select the matching quantity as represented in dots from the two choices provided. Overall accuracy was only 56%, suggesting that performance was at floor and participants were guessing. This task was omitted from further analyses.

Results

Data Preparation and Descriptive Statistics

Prior to conducting analyses, the data were screened for normality and extreme values. For most tasks with RTs, less than 5% of trials were spoiled. Trials were considered spoiled when RTs were missing or less than 250ms. In some cases, trials were considered spoiled because the experimenter entered a missing or invalid code (e.g., 999); this could have been due to events like a question from the participant, a distraction, or a mistriggered voice key). For some tasks, the experimenter entered participant responses, and some answers reflected experimenter error (e.g., the task was to name the number 7, and the "74" was entered). In most cases these trials were simply discarded (i.e., not counted as errors in accuracy calculations); overall the amount of data lost was negligible—with two main exceptions. In the nonsymbolic quantity identification tasks, stimuli were presented for a finite amount of time (2000ms). If the participant did not respond during that time, the stimuli presentation software advanced to the text response box, and the voice key RT was not collected. Nonsymbolic identification trials in the medium and large ranges were missing 31.5% and 42% of RTs respectively. RTs were not used as an index of performance on the large nonsymbolic identification task. However, RT was used to calculate *P* (see below) for nonsymbolic identification in the medium range. This was done because this measure was the closest measure to counting tasks employed in other research. There were 20 trials in the medium range, thus even with a loss of 31.5% of trials on average, RTs could still be reliable (and this measure did in fact have a good split-half reliability; see Table 3.5). Otherwise, very few extreme scores were identified overall. Three participants had extreme scores on the symbolic number line tasks. One participant had an extreme Go/No Go score. These variables were coded as missing, and the rest of the participants' data were retained, as MPlus estimates missing data. Some measures were lacking data for some participants; this was sometimes due to equipment failure or time constraints while testing. This is reflected in *N*s reported in Tables 3.5 and 3.6.

Descriptive statistics for all measures with response times (RTs) are shown in Table 3.5; descriptive statistics for the other measures are shown in Table 3.6. (Task scoring and reliabilities are discussed below). Performance on the symbolic versus nonsymbolic quantitative tasks was compared, and results are shown in Table 3.7. In general, the nonsymbolic tasks tended to have lower accuracies and higher RTs than the symbolic versions of the same tasks; almost all of these differences were significant. This suggests that nonsymbolic tasks were more challenging. It is reasonable to assume that the symbolic and nonsymbolic tasks were accessing different numerical representations. It is interesting to note that RT and accuracy for the same task were often negatively correlated, suggesting some speed-accuracy trade-off. The pattern within the ordering task varied from symbolic to nonsymbolic formats. In the symbolic format, when participants were faster, they also tended to be more accurate. In contrast, on the nonsymbolic ordering task, the faster participants responded on average, the less accurate they were. Perhaps when participants were unsure they were more likely to quickly guess.

Table 3.5.

Study 3 Descriptive Statistics and Reliabilities for RT Tasks

		Median RT (ms)					*P*				
Measure	*N*	Min	Max	*M*	*SD*	Acc.	Min	Max	*M*	*SD*	Rel.
Symbolic Quantitative											
Identification small[1]	153	375	652	510	60	--	--	--	--	--	.95
Identification medium[1]	153	381	658	505	58	--	--	--	--	--	.96
Identification large[1]	153	440	886	576	90	--	--	--	--	--	.97
Comparison small	151	358	745	476	73	98.5	358	965	490	83	.80
Comparison medium	151	371	824	529	89	91.8	395	1099	614	114	.69
Comparison large	151	452	996	677	111	93.8	515	1111	757	118	.94
Ordering small-medium	152	518	1432	921	194	89.8	577	2090	1113	299	.94
Ordering large	152	614	1549	1085	185	89.9	689	2370	1311	278	.92
Nonsymbolic Quantitative											
Identification small[1]	151	519	1277	702	125	100	--	--	--	--	.78
Identification medium	150	649	1902	1320	269	4.7[2]	671	1944	1378	267	.82
Comparison small	150	388	830	547	87	97.5	392	830	573	93	.73
Comparison medium	150	445	1163	737	169	90.4	505	1710	873	210	.63
Comparison large[3]	150	424	1392	753	192	77.1	--	--	--	--	--[4]
Ordering small-medium[3]	151	448	1617	1073	251	76.4	--	--	--	--	--[4]
Ordering large[3]	151	360	1512	991	264	69.7	--	--	--	--	--[4]
Executive Function											
Choice RT	150	335	560	414	48	95.8	347	653	447	56	.81
Black White Stroop[1]	147	461	1004	461	103	--	--	--	--	--	.93
Mathematical Outcome											
Speeded Addition	149	669	2305	1086	276	95.8	702	2603	1184	358	.89

Note: [1]No *P* because accuracy 100% or not available. [2]Mean percent absolute error. [3]No *P* because accuracy < 85%. [4]Reliability in Table 3.7. ms = milliseconds. Acc. = accuracy. Rel. = Split-half reliability.

Table 3.6.

Study 3 Descriptive Statistics and Reliabilities for Non-RT Measures

		Scores				
Measure	*N*	*Min*	*Max*	*M*	*SD*	Rel.
Visual Spatial Span[1]	147	3	22	16.2	3.1	.77[a]
Backward Digit Span[1]	150	2	12	7.0	2.6	.82[b]
Go/No Go Task						
Accuracy Go trials (%)	144	93	100	98.7	1.58	
Percent false alarms	144	0	75	32.6	19.2	
d'	144	-4.51	2.33	.04	1.41	
Rapid ID Symbolic[2]	153	2.00	4.77	3.32	.66	.86[c]
Nonsymbolic quantitative						
Rapid ID small[2]	153	1.63	3.64	2.60	.39	.85[c]
ID large[3]	150	11	42	25.4	6.8	.97[b]
Comparison large[4]	150	55	90	77.1	5.9	.80[b]
Ordering small-medium[4]	151	46	93	76.4	9.8	.83[b]
Ordering large[4]	151	33	90	69.7	8.9	.75[b]
Arithmetic Fluency (CFT)						.91[a]
Addition[1]	153	5	38	15.9	6.6	
Subtraction[1]	153	1	30	12.0	5.4	
Multiplication[1]	152	0	24	7.1	5.6	
Calc. Knowledge (BMA) [1]	153	2	11	7.2	2.2	.69
Symbolic number line						.74[a]
Linear slope (*b*)	150	.59	1.1	.95	.07	
Linear R^2	150	.66	1.0	.97	.04	
M % absolute error	150	2	15	4.7	2.1	
Nonsymbolic number line						.73[a]
Linear slope (*b*)	153	.08	1.1	.81	.20	
Linear R^2	153	.03	.97	.81	.16	
M % absolute error	153	4	43	12.4	.06	

Note: [1]Total correct. [2]Items-per-second. [3]Mean percent absolute error. [4]Accuracy (%). Rel. = Reliability. [a]Internal (Cronbach's α); [b]Split-Half; [c]Parallel-form. [c]Internal (Cronbach's α) as published. ID = Identification. Calc. = Calculation. CFT = Calculation Fluency Test.

Table 3.7.

Symbolic and Nonsymbolic Quantitative Comparisons: RT and Accuracy

	Mean Accuracy (%)			RT (ms)			*r* RT and accuracy	
	S	NS	Δ	S	NS	Δ	S	NS
Identification								
Small	N/A	0.0[a]	--	510	702	192**	--	--
Medium	N/A	4.7[a]	--	505	1320	815**	--	--
Large	N/A	25.4[a]	--	576	N/A	--	--	--
Comparison								
Small	98.5	97.5	1.0*	477	545	68**	-.08	-.19*
Medium	91.7	90.5	1.2	529	735	206**	-.21*	-.22*
Large	93.8	77.1	16.7**	676	753	77**	-.32**	-.31**
Ordering								
Sm-Medium	90.0	76.4	13.6**	920	1073	153**	.21*	-.44**
Large	89.7	69.7	20.0**	1083	1003	80**	.13	-.51**

Note. $^{*}p \leq .05$, $^{**}p \leq .001$. RT = response time, S = Symbolic, NS = Nonsymbolic, N/A = not available. Δ = difference between symbolic and nonsymbolic. ms = milliseconds. *r* = Pearson bivariate correlation. [a]Participants had to name the number of dots, and thus were not expected to be exactly correct, hence the use of absolute error. Percent absolute error was calculated by dividing by the scale, which was 9 for small and medium (administered as one task; participants were told they would see 1 to 9 dots) and 90 for large (participants were told they would see between 10 and 100 dots).

Variable scoring. See Table 3.8 for a summary of tasks and their final performance measures/scores (described here). For the nonsymbolic and symbolic rapid identification tasks, items-per-second scores were calculated for each form by subtracting the number of errors from the number of items, then dividing by the total time taken to name the stimuli (i.e., items-per-second = [[24 - errors]/time]).

For computerized tasks, when response times (RTs) and accuracy were both available, and accuracy was greater than 85%, *P* was calculated as the measure of performance (Lyons et al., 2014). Unless otherwise specified, *P* was calculated with the following formula: $P = mdn\mathrm{RT}(1 + (2*ER))$. ER is the error rate; the error rate was multiplied by 2 since most RT tasks required binary forced-choice responses (ER = .5 represents chance performance). *P* stands for performance and represents the median RT in milliseconds with time added on to penalize for errors. If accuracy is perfect, *P* and median RT are the same. This performance indicator was selected for two reasons: 1) Rather than having two separate error and RT variables, or selecting one over the other, *P* combined the performance aspects of both to create a single variable for use in later analyses. 2) Creating a *P* score was most consistent with the scoring of the rapid identification tasks. These tasks were scored by creating an items-per-second score which were corrected by the number of errors (if any were made). Thus, both *P* and the items-per-second score incorporate response time and error rates into one measure. *P* was not calculated for tasks with accuracies less than 85% because a) the *P* scores no longer closely resembled RT scores, and b) accuracy contained enough variability to serve as a suitable independent or dependent variable.

A modified *P* was calculated as the nonsymbolic identification medium (5 to 9) performance indicator. This task involved naming/identifying quantities. Stimuli presentation time was limited, so, at times participants were not exactly correct. Percent absolute error (PAE) was calculated by taking the absolute error and dividing by the scale, which was 9 (stimuli were 1 to 9 dots in the small-medium block). That is, PAE for each trial was calculated with the formula: PAE = (|stimulus - response| / 9). The mean PAE across all nonsymbolic identification trials in the medium range was calculated per participant using the aggregate function in SPSS (as was median RT). *P* was then calculated with the following formula: $P = mdn\mathrm{RT}^*(1+M\ \mathrm{PAE})$.

A large proportion of trials in the large nonsymbolic identification block lacked valid RTs due to participants not responding during the 2000ms stimuli-presentation window (i.e., they timed out), and thus an RT performance measure could not be calculated with confidence. Instead percent absolute error (PAE) on each trial was calculated by taking the absolute error and dividing by the scale (90; stimuli were 10 to 100 dots). That is, PAE was calculated with the following formula: PAE = (|stimulus - response| / 90). The mean PAE was calculated with the aggregate function in SPSS.

P was not calculated for the nonsymbolic identification of small quantities because accuracy was perfect, thus median RT was the measure of performance. *P* was also not calculated for symbolic identification (small, medium, and large) and Black-White Stroop. In these tasks the RT was recorded with a voice-key, after which the experimenter entered the participant's response. Errors on these tasks were extremely rare, and were as likely be experimenter as participant error, thus median RT (without an error correction) was the performance measure for these variables.

Table 3.8. *Study 3 Performance Indicators*

	Nonsymbolic	Symbolic
Quantitative Skills		
Identification		
Rapid Identification Small	*M* items-per-second	*M* items-per-second
Small	RT	RT
Medium	*P*	RT
Large	*M* Abs. Error	RT
Comparison		
Small	*P*	*P*
Medium	*P*	*P*
Large	Accuracy	*P*
Ordering		
Small-Medium	Accuracy	*P*
Large	Accuracy	*P*
Executive Function (5 tasks below as indicators for a latent factor)		
Choice RT (attention and processing speed)		*P*
Black White Stroop (inhibitory control)		RT
Go/No Go (inhibitory control)		*d'*
Visual Spatial span (visual spatial working memory)		Total correct spans
Backward Digit Span (central executive and phonological loop)		Total correct spans
Experiential/Demographic Factors		
Perceived Math Competence (5 indicators; score 1-7)		Latent factor
Age		Age in years
Education Location (Asian-educated =2, Elsewhere = 1)		Dichotomous variable
Dependent Measures		
Arithmetic Fluency (3 subscales; total correct)		Latent factor
Speeded Addition		*P*
Calculation knowledge (Brief Math Assessment)		Total correct
Nonsymbolic Number Line (3 indicators: slope, R^2, & abs. error)		Latent factor
Symbolic Number Line (3 indicators: slope, R^2, & abs. error)		Latent factor

Note: RT = Response time. Abs. = Absolute. *P* = median RT corrected by error rate.

Reliabilities. Reliabilities are presented along with descriptive statistics in Tables 3.5 and 3.6. For many tasks, trials were split into two balanced halves and split-half reliability was calculated. Note that reliability was calculated with each variables' performance measure. For example, if *P* was the performance measure, *P* was calculated for each half and then split-half reliability was calculated. In some cases internal scale reliability was calculated; this was done with the three calculation fluency subtests, the perceived math competence items, and the indicators of symbolic and nonsymbolic number line performance. This was also done with the visual spatial span task. The task was scored as three separate subscores, which included one trial from each span size (i.e., each subscore included one trial each of spans 2, 3, 4, 5, 6, 7, 8, and 9). The three resulting scores were then entered as items of a scale to determine the observed internal reliability coefficient.

Data transformation. Prior to correlational analyses, the data were transformed so that all variables would share a positive valence (i.e., high values represent better performance). *P,* median RT, and absolute error scores were multiplied by -1, and absolute error scores were reverse scored.

Overview of Data Analysis

The nonsymbolic quantitative measures were expected to group together by range (small, medium, and large)—representing the subitizing, counting, and approximate number systems, respectively. Conversely, symbolic measures were believed to tap a single underlying representation (rather than three). So, while the symbolic measures were expected to group together by range, failure to do so would not have the same theoretical implications as if it occurred with the nonsymbolic tasks. This mentioned as rationale for why nonsymbolic and symbolic quantitative measures might be treated differently in the data analyses described below.

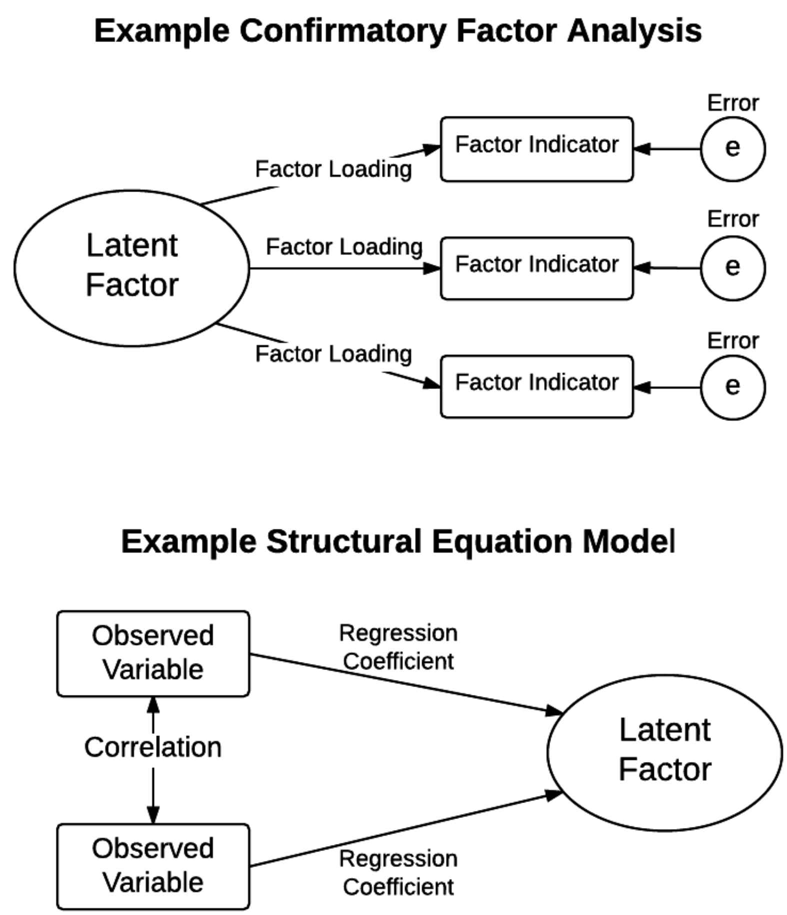
Five different mathematical/numerical measures served as outcomes in the present study: Arithmetic fluency, Speeded Addition, Calculation Knowledge (BMA), Symbolic Number Line, and Nonsymbolic Number Line. The Arithmetic Fluency and the Symbolic and Nonsymbolic Number Line outcomes were latent factors; Speeded Addition and Calculation Knowledge were observed variables.

Data analysis was conducted in five steps as described below. Steps 3 through 5 were conducted separately for each mathematical outcome. Step 1 entailed examining the strength of correlations among intended factor indicators, observed predictors, and mathematical outcomes. Confirmatory factor analyses (CFAs) were conducted in Step 2 to achieve three goals: 1) To test the hypothesis that nonsymbolic tasks within the same ranges (small, medium, and large) tap the same underlying constructs—namely subitizing, counting, and approximate number systems (H8; this structure was also tested for symbolic quantitative measures); 2) To select nonsymbolic quantitative variables, and to create latent symbolic quantitative, executive function, and perceived math competence factors to act as predictors of mathematical outcomes in structural equation models (SEMs); and 3) To examine the fit of the measurement model (all of the factors specified at the same time). In Step 3, correlations among factors and observed variables are presented. In Step 3a the correlations common among all models are discussed. In Step 3b, correlations among predictors and individual mathematical outcomes are discussed (done separately for each outcome). If the mathematical outcome was a latent factor, it was added to the measurement model at Step 3. In Step 4, a structural equation model (SEM) was conducted with only the domain-specific quantitative skills included as predictors of the math outcome (called the “quantitative-only” SEMs). Quantitative skills were also examined in relation to each other. Finally, in Step 5, executive function (domain-general) and experiential constructs (age, education location, and PMC) were added to the existing SEM; these were called the “Full SEMS”.

CFAs and SEMs were run in the statistical program MPlus (Muthén & Muthén, 1998-2012) with full information maximum likelihood (FIML) estimation, which estimates missing data (Schreiber et al., 2006). Factor scales were standardized by setting factor means to 0, factor variances to 1; and specifying the first factor loading to be free (rather than fixed to 1). CFAs and SEMs diagrams represent model parameters in a conventional way. These conventions are described in the text here, and summarized in Figure 3.1. Ovals represent latent factors. Arrows from a factor (oval) to a rectangle represent factor loadings; the rectangles represent the factor indicators. The circles with arrows pointing to indicators represent indicator error. Other rectangles in a path or SEM analysis represent observed variables. Single-headed arrows between variables represent regression paths. Double-headed arrows represent correlations; note that in the present research these arrows could be curved or straight. The values shown in the figures are the standardized coefficients (e.g., factor loadings, correlations, and regression coefficients).

Several model fit indicators were used, consistent with common usage in the literature (Hu & Bentler, 1999; Schreiber, et al., 2006), to evaluate whether models were a good fit to the data. These included the following: *Chi-square test of model fit* (χ^2) will be non-significant for models that fit well; significant values indicate model misfit. However, model fit can still be considered acceptable when Chi-square is significant if the ratio of χ^2 to *df* is less than or equal to 2 or 3 (Shreiber et al., 2006). The *Comparative Fit* and *Tucker-Lewis* indices (CFI and TLI) should equal or exceed .95. The *standardized root mean square residual* (SRMR) should be less than or equal to .08, and the *root mean square error of approximation* (RMSEA) between .06 and .08. *RMSEA 90% confidence intervals* were also reported (lower confidence interval =LCI; upper confidence interval = UCI).

Figure 3.1. Summary of CFA and SEM figure components.



Step 1: Correlations Among Observed Variables

Table 3.9 displays correlations among executive function, perceived math competence, and mathematical variables. Correlations among variables intended to load onto the same factors are shown in boxes. Executive function measures (constructs in brackets) were: Choice RT (attention and processing speed), Black White Stroop (inhibitory control), Go/No Go (inhibitory control), Visual Spatial Span (visual spatial working memory) and Backward Digit Span (central executive and phonological loop). The observed moderate correlations among these measures support the assumption that they tap a shared domain-general cognitive construct. Arithmetic fluency indicators (addition, subtraction, and multiplication) were highly correlated, as were PMC (perceived math competence) items, supporting the notion that these items tap the same underlying constructs. Correlations across factors are discussed in Step 2.

Correlations between education location, age, and each of executive function, PMC, and mathematical variables appear in Table 3.10. Education location and age did not correlate with executive function measures. Compared to those educated elsewhere, Asian-educated participants felt slightly more competent in their basic math skills and performed better on the mathematical computation measures, but not on the number line tasks. Relative to younger adults, older adults self-reported lower levels of perceived math competence, but also performed significantly better on arithmetic fluency measures. The pattern of correlations supports the inclusion of education location and age as predictors of mathematical outcomes in SEM analyses.

Table 3.9.

Correlations: Executive Function, PMC, and Mathematical Outcomes.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.
1. Choice RT	--																			
2. BW Stroop	**.44***	--																		
3. Go/No Go	**.33***	**.44***	--																	
4. VSp span	**.27***	**.25***	**.38***	--																
5. BDig span	**.28***	**.24***	**.19***	**.35***	--															
6. PMC-1	**.18***	.10	.11	**.21***	**.20***	--														
7. PMC-2	.15	.09	.19	**.24***	.11	**.58***	--													
8. PMC-3	**.19***	.00	.15	**.30***	**.20***	**.51***	**.59***	--												
9. PMC-4	.09	.03	.12	**.20***	.09	**.46***	**.43***	**.71***	--											
10. PMC-5	.00	.01	.12	**.16***	.04	**.34***	**.50***	**.49***	**.46***	--										
11. CFT Add.	**.31***	**.19***	.08	.15	**.27***	**.32***	.16	**.28***	.15	.03	--									
12. CFT Sub.	**.25***	**.17***	.14	**.18***	**.31***	**.41***	**.23***	**.32***	**.21***	.07	**.80***	--								
13. CFT Mult.	**.20***	.09	.07	.10	**.23***	**.34***	.14	**.25***	.13	-.05	**.77***	**.75***	--							
14. Sp.Add.	**.33***	**.19***	.13	**.23***	**.35***	**.46***	**.30***	**.40***	**.29***	.08	**.60***	**.65***	**.59***	--						
15. Calc.Know.	**.12**	**.19***	.10	**.26***	.12	**.40***	**.38***	**.34***	**.26***	.09	**.37***	**.41***	**.41***	**.49***	--					
16. S NL *b*	.12	.03	**.24***	**.24***	.11	.07	-.03	.12	**.18***	-.03	.04	.05	.02	**.18***	.06	--				
17. S NL R^2	**.21***	.10	**.24***	**.21***	**.21***	.13	.07	**.26***	**.20***	-.05	**.22***	**.22***	**.18***	**.38***	**.18***	**.63***	--			
18. S NL AE	**.28***	**.18***	**.33***	**.32***	**.23***	**.18***	.13	**.27***	**.19***	.00	**.23***	**.24***	**.18***	**.45***	**.26***	**.63***	**.81***	--		
19. NSNL *b*	**.21***	**.22***	.13	**.18***	.15	-.12	-.09	-.01	-.01	-.04	.08	.05	.01	.14	.09	**.41***	**.27***	**.38***	--	
20. NS NL R^2	**.24***	**.21***	.12	**.19***	.15	-.11	-.08	.00	.00	-.12	.10	.04	.01	.11	.10	**.38***	**.40***	**.47***	**.81***	--
21. NS NL AE	**.24***	**.19***	.01	**.19***	**.16***	-.01	.02	.06	.08	-.04	.12	.11	.08	**.20***	**.23***	**.39***	**.33***	**.43***	**.88***	**.79***

Note. $^*p \leq .05$; BW = Black White. VSp = Visual spatial. BDig = Backward digit. PMC = Perceived math competence. CFT = Calculation fluency test. Add. = Addition. Sub. = Subtraction. Mult. = Multiplication. Sp.Add. = Speeded addition. Calc.Know. = Calculation knowledge (BMA). S = Symbolic. NL = Number line. NS = Nonsymbolic. *b* = slope. AE = absolute error (reverse coded).

Table 3.10.

Study 3 Correlations with Education Location and Age

Measure	Education Location	Age
Choice RT	-.02	-.12
Black White Stroop	-.10	-.11
Go/No Go	-.08	-.02
Visual Spatial Span	.00	-.11
Backward Digit Span	-.09	.11
PMC-1	**.17***	**-.22***
PMC-2	.06	**-.33****
PMC-3	.06	-.07
PMC-4	.07	-.15
PMC-5	-.04	**-.22****
CFT Addition	**.32****	**.24****
CFT Subtraction	**.40****	**.23****
CFT Multiplication	**.36****	**.21****
Speeded Addition	**.21***	.08
Calc. Knowledge (BMA)	**.23***	.03
Symbolic Number Line		
b	.02	-.08
R^2	.07	.05
Absolute Error	.01	.05
Nonsymbolic Number Line		
b	-.02	.08
R^2	-.01	.10
Absolute Error	.00	.05

Note. $^{*}p \leq .05$; $^{**}p \leq .01$. Absolute error is reverse coded.
Education location is coded 2 = Educated in Asia; 1 = Educated elsewhere.
PMC = Perceived math competence. CFT = Calculation fluency test.
Calc. = Calculation. BMA = Brief Math Assessment-III.

Quantitative correlations. Quantitative measures were expected to group together based on format (symbolic or nonsymbolic) and range (1 to 3, 5 to 9, and 10 to 100) to create symbolic small, medium, and large factors, and nonsymbolic small, medium, and large factors. Correlations among quantitative measures are shown in Table 3.11; correlations among measures expected to share a common factor are shown in boxes.

Nonsymbolic quantitative. Small-range nonsymbolic variables were weakly correlated. Medium nonsymbolic identification and comparison tasks were correlated, however the small-medium ordering measure did not correlate with other medium measures. Large nonsymbolic comparison and ordering were moderately correlated, but large identification did not correlate with either of them. In contrast to the low correlations within size ranges, high correlations were observed among variables derived from similar tasks (i.e., identification, comparison, or ordering), suggesting that task demands might account for more shared variance than range.

Symbolic quantitative. Correlations among symbolic small, medium, and large factor indicators were small to moderate, with some exceptions. Medium comparison and small-medium ordering were highly correlated, as were large comparison and ordering measures. As was observed with the nonsymbolic measures, variables sharing task demands were more highly correlated than variables within the same size range, suggesting that factors might be better formed by grouping variables by task (identification, comparison, and ordering) than by range (small, medium, and large).

Table 3.11.

Study 3 Quantitative Tasks: Correlations by Range (small, medium, and large).

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.
Nonsymbolic																		
1. R ID Sm	--																	
2. ID Sm	**.27***	--																
3. Comp Sm	**.26***	**.26***	--															
4. ID Med	**.34***	**.47***	**.37***	--														
5. Comp Med	**.25***	**.33***	**.65***	**.36***	--													
6. Ord. Sm-Med	.09	.15	**.16***	.10	.00	--												
7. ID Lg	.13	.14	**.16***	**.26***	.11	.12	--											
8. Comp Lg	.06	.03	.10	.01	.05	**.53***	.14	--										
9. Ord. Lg	**.25***	.12	.13	.13	.04	**.64***	.12	**.49***	--									
Symbolic																		
10. R ID Sm	**.75***	**.25***	**.17***	**.26***	.13	.02	.14	.01	**.18***	--								
11. ID Sm	**.34***	**.41***	**.18***	**.42***	**.23***	.12	**.25***	.01	.12	**.32***	--							
12. Comp Sm	**.37***	.14	**.39***	**.33***	**.35***	**.25***	**.21***	**.26***	**.24***	**.19***	**.27***	--						
13. ID Med	**.34***	**.42***	**.21***	**.41***	**.22***	.15	**.29***	.05	.14	**.31***	**.94***	**.34***	--					
14. Comp Med	**.23***	.13	**.27***	**.25***	**.28***	**.34***	**.21***	**.35***	**.27***	.06	**.19***	**.43***	**.24***	--				
15. Ord. Sm-Med	**.41***	**.19***	**.44***	**.39***	**.31***	**.43***	**.22***	**.25***	**.28***	**.26***	**.18***	**.54***	**.24***	**.61***	--			
16. ID Lg	**.32***	**.37***	**.20***	**.31***	.14	.05	**.29***	-.05	.01	**.40***	**.71***	**.22***	**.75***	.09	**.17***	--		
17. Comp Lg	**.45***	**.23***	**.56***	**.37***	**.47***	**.27***	**.24***	**.30***	**.27***	**.29***	**.31***	**.62***	**.37***	**.65***	**.77***	**.27***	--	
18. Ord. Lg	**.40***	**.21***	**.49***	**.36***	**.38***	**.46***	**.18***	**.33***	**.38***	**.26***	**.20***	**.50***	**.26***	**.58***	**.85***	.14	**.77***	--
Ed Location	.07	-.10	.02	-.05	.04	.00	-.03	.02	.03	-.14	**-.24***	.16	**-.25***	.12	.15	**-.49***	.12	.12
Age	.01	.14	-.11	.02	-.09	**.16***	.07	.15	.06	.03	.09	-.02	.09	.04	.09	.15	.00	.02

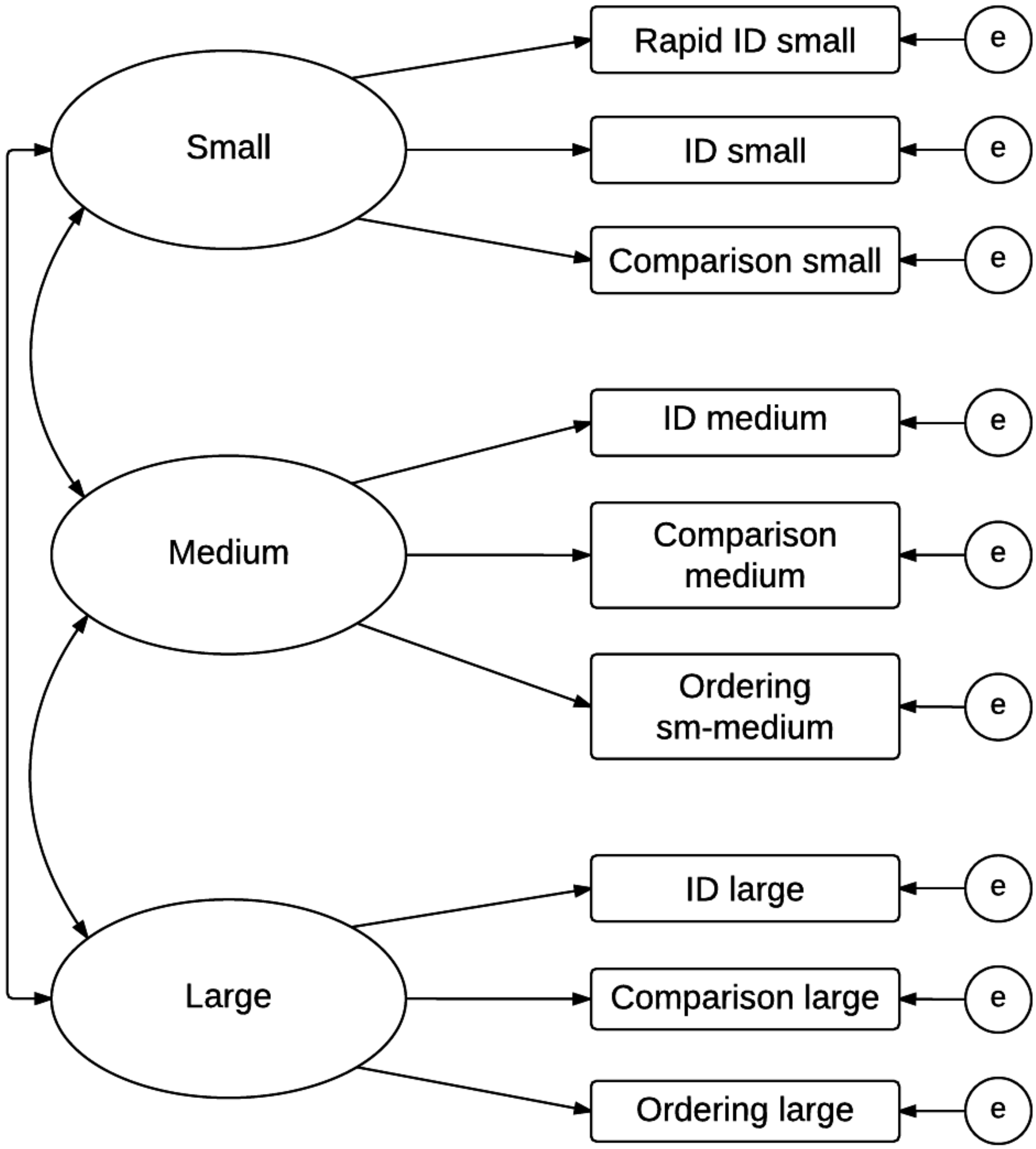
Note: $^*p \leq .05$. R = Rapid. ID = Identification. Comp. = Comparison. Ord. = Ordering. Sm = Small (1 – 3). Med = Medium (5-9). Sm-Med = Small-Medium (1 – 9). Lg = Large (10 – 100). CFT = Calculation Fluency Test. Ed = education.

Step 2: CFAs, Variable Selection, and Test of Measurement Model

Quantitative CFAs. Quantitative measures were expected to group together based on format (symbolic or nonsymbolic) and range (1 to 3, 5 to 9, and 10 to 100) to create symbolic small, medium, and large factors, and nonsymbolic small, medium, and large factors. This hypothesized factor structure is shown in Figure 3.2. CFAs were conducted separately for nonsymbolic and symbolic quantitative measures.

Nonsymbolic quantitative CFA. Nonsymbolic identification, comparison, and ordering tasks from the same range (small, medium, large) were hypothesized to tap the same underlying constructs—namely, subitizing, counting, and ANS, respectively (H8). These constructs were of theoretical importance in the present research, thus this hypothesized factor structure was tested, despite the low correlations observed among nonsymbolic variables of the same range because CFA might still capture shared variance. When the nonsymbolic CFA by range (small, medium, and large) was run, an error message was returned indicating the covariance matrix was not positive definite, and thus results were not valid. The model was re-specified (e.g., variables were removed) and attempted a few more times but the covariance matrix remained not positive definite (i.e., H8 was not supported). As described above, measures from similar tasks seemed to share more variance than variables within the same range, however a nonsymbolic CFA structured by task was not conducted. Doing so would have collapsed nonsymbolic quantitative measures across the small, medium, and large ranges, and thus would not be consistent with the goal of simultaneously examining subitizing, counting and ANS skills as predictors of mathematical outcomes. Instead, specific variables were selected to represent subitizing, counting, and ANS performance for SEM analyses.

Figure 3.2. Expected factor structure for symbolic and nonsymbolic quantitative tasks.



Note: ID = Identification. Small refers to quantities 1 to 3, medium to quantities 5 to 9, and large to quantities 10 to 100. sm-medium = small medium (quantities 1 to 9).

Subitizing, counting and ANS measures were selected based on their similarity with tasks used in past research to tap the same constructs. Nonsymbolic rapid identification of small quantities (1 to 3) was selected as the subitizing measure; this measure was also observed to be reliable and valid in Studies 1 and 2. Medium-range nonsymbolic identification (5 to 9) was selected as the counting measure, and large-range nonsymbolic comparison was selected as the index of the ANS. The final SEM included symbolic ordering (details below), therefore the large nonsymbolic ordering measure was also included in an attempt to make the symbolic and nonsymbolic quantitative predictors as similar as possible.

Symbolic quantitative CFAs. Refer to Figure 3.2 for the expected factor structure and to Table 3.11 for correlations among symbolic quantitative measures. The initial CFA resulted in a not positive definite covariance matrix. The output also provided a warning about the medium identification variable. One apparent issue in the MPlus output was an estimated correlation between small and medium symbolic factors greater than 1. This could have been due, in part, to the .95 correlation between medium and small identification tasks.

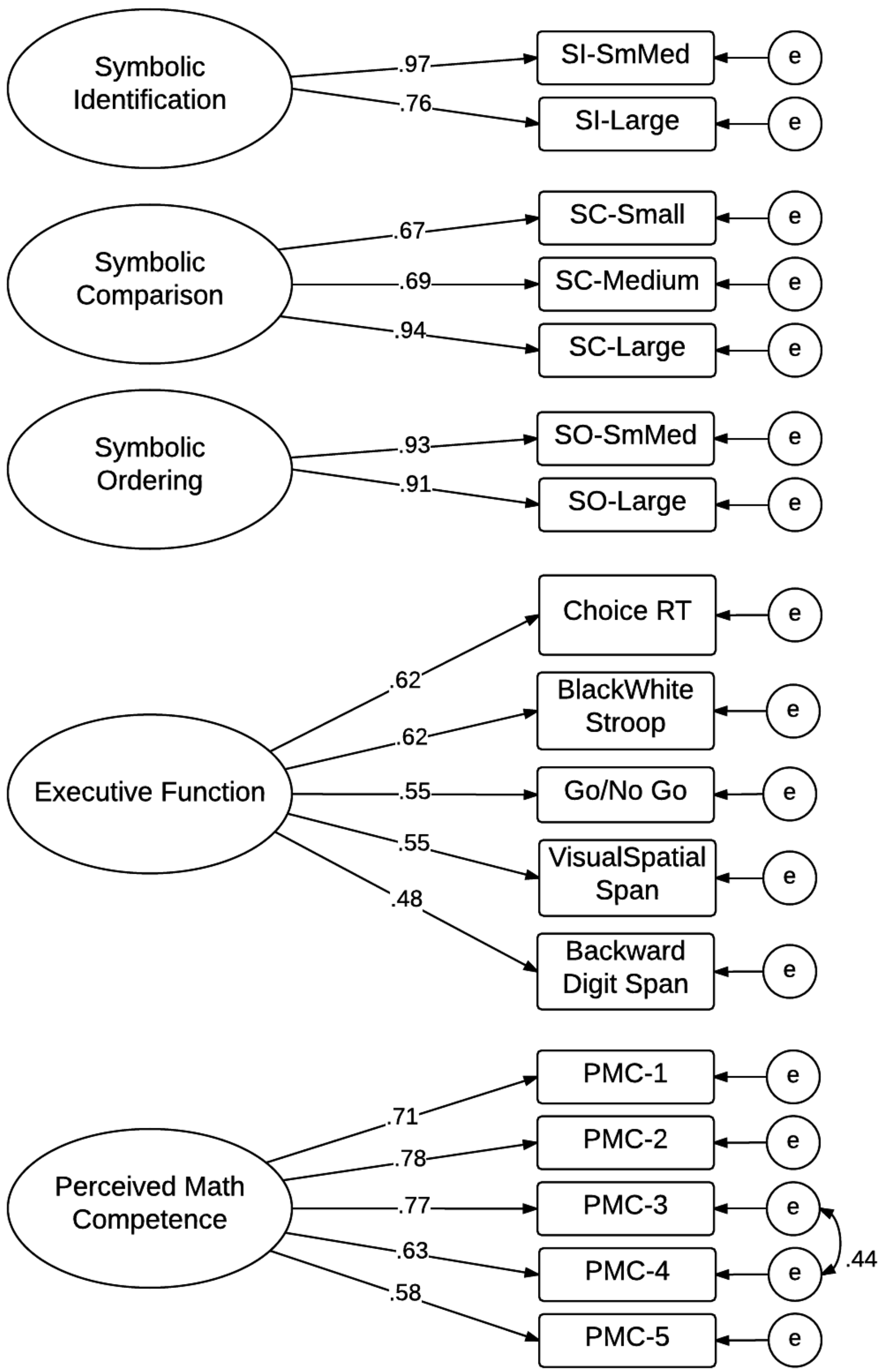
Given the pattern of correlations described above, another symbolic CFA was conducted with measures grouped by task (identification, comparison, and ordering) rather than range (small, medium, and large). The small and medium identification measures were rescored as one variable (small-medium identification), given their correlation of .95. The symbolic rapid identification of small quantities was highly correlated with the analogous nonsymbolic task chosen as the measure of subitizing

(r = .75). Due to concerns about multicollinearity, the symbolic rapid identification task was excluded from the symbolic identification factor and from further analyses.

Rather than running the symbolic quantitative CFA-by-task on its own, the larger measurement model was tested: Symbolic identification, comparison, and ordering factors were specified, as were executive function, and perceived math competence (PMC) factors. These factors were allowed to correlate. Refer to Figure 3.3 for factor indicators and standardized factor loadings. Estimated correlations among factors are provided in Table 3.12 rather than shown with the measurement model. See Table 3.13 for model fit indices (Measurement Model v1). The initial fit of the measurement model was a little less than ideal. The addition of a correlation between the error terms for PMC items 3 and 4 was recommended. These items asked about nervousness and avoidance associated with math; it seemed reasonable for they would share variance related to math anxiety not captured by the PMC factor, therefore this correlation was added. Model fit improved and was accepted as adequate (see Measurement Model v2 in Table 3.13).

Symbolic identification, comparison, and ordering factor loadings were close to or above .70, which was ideal. Executive function factor loadings were a bit lower, but above .40, and thus were considered adequate. PMC factor loadings were also within a reasonable range. In sum the measurement model seemed to be a suitable fit to the data and produced meaningful factors for use in SEM analyses.

Figure 3.3. Measurement model: Symbolic quantitative, executive function, PMC factors.



Note: Standardized solution shown. Factors were set to be correlated; correlations are not shown in the figure, but rather presented in Table 3.13.

Table 3.12.

Study 3 Correlations: Observed Variables and Latent Factors

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. PMC[a]	--								
2. Executive Function[a]	**.32****	--							
NS quantitative									
3. Subitizing[b] (RapidID sm)	**.24***	**.65****	--						
4. Counting[b] (ID medium)	.18	**.51****	**.34****	--					
5. ANS[b] (Comp lg)	.05	**.28***	.07	.02	--				
6. Ordering large[b]	.13	**.43****	**.25****	.13	**.49****	--			
Symbolic quantitative									
7. Identification[a]	.10	**.60****	**.37****	**.43****	.02	.13	--		
8. Comparison [a]	**.38****	**.70****	**.47****	**.40****	**.35****	**.30****	**.39****	--	
9. Ordering [a]	**.44****	**.61****	**.43****	**.40****	**.30****	**.36****	**.25***	**.89****	--
Math outcomes									
Arithmetic Fluency[a]	**.35****	**.38****	**.46****	**.27****	.12	**.19***	.12	**.53****	**.65****
Speeded Addition[b]	**.47****	**.44****	**.37****	**.27****	.12	**.17***	**.20***	**.68****	**.68****
Calc.Knowledge[b] (BMA)	**.46****	**.21***	.15	.09	.13	**.20***	.03	**.41****	**.39****
Symbolic Number Line[a]	**.21***	**.45****	**.17***	**.20***	**.35****	**.36***	.11	**.48****	**.50****
Nonsymbolic Number Line[a]	-.07	**.35****	.10	**.18***	**.41****	**.40****	.13	**.26***	**.25***

Note. $^{*}p \leq .05$, $^{**}p \leq .01$. [a]Latent factors. [b]Observed variables. NS = Nonsymbolic. PMC = Perceived math competence. ANS = Approximate number system. ID = identification. Comp = comparison. Calc. = Calculation. BMA= Brief math assessment.

Table 3.13.

Study 3 Summary of Model Fit of Structural Equation Models

	df	χ^2	*p* value	Ratio χ^2/df	CFI	TLI	SRMR	RMSEA	RMSEA 90% CI Lower	RMSEA 90% CI Upper
Cutoff criteria			*n.s.*	≤ 2 or 3	≥ .95	≥ .95	≤ .08	<.06 to .08		
Model										
Measurement Model v1 (no DVs)	109	178.7	< .001	1.6	.942	.927	.065	.065	.047	.081
Measurement Model v2 (no DVs)	108	165.8	< .001	1.5	.952	.939	.064	.059	.040	.076
Arithmetic Fluency										
Measurement Model	154	223.4	< .001	1.5	.956	.946	.066	.054	.038	.069
Quantitative Only Model	66	92.8	.016	1.4	.978	.969	.063	.052	.023	.075
Final SEM	276	452.3	< .001	1.6	.909	.893	.078	.065	.054	.075
Speeded Addition										
Quantitative Only Model	43	59.4	.049	1.4	.982	.972	.059	.050	.004	.079
Final SEM	229	407.6	< .001	1.8	.888	.867	.077	.071	.060	.083
Calculation Knowledge (BMA)										
Quantitative Only Model	44	60.5	.050	1.4	.980	.971	.061	.049	.001	.078
Final SEM	230	397.3	< .001	1.7	.891	.870	.077	.069	.057	.080
Symbolic Number Line										
Measurement Model	154	225.3	< .001	1.5	.952	.941	.064	.055	.039	.070
Quantitative Only Model	66	90.5	.02	1.4	.978	.970	.060	.049	.019	.073
Final SEM	276	462.6	< .001	1.7	.896	.878	.077	.066	.056	.077
Nonsymbolic Number Line										
Measurement Model	154	233.8	< .001	1.5	.951	.940	.060	.058	.042	.073
Quantitative Only Model	66	96.3	.009	1.5	.976	.967	.061	.055	.028	.077
Final SEM	276	466.0	< .001	1.7	.902	.886	.074	.067	.056	.077

Step 3a: Correlations among factors and observed variables

Correlations among the mathematical outcomes (latent factors and observed variables) were generated in MPlus and are shown in Table 3.12. Estimated correlations among predictors (latent factors and observed variables) and mathematical outcomes were also generated by creating models in MPlus and requesting the correlations; this was done separately for each mathematical outcome. Correlations varied slightly by model, but never more than .01. This was deemed to be a negligible amount of variation, and one table of correlations among predictors and mathematical outcomes is presented, rather than multiple. Correlations between predictors and each mathematical outcome are described separately in Step 3b of the analyses for each outcome.

Table 3.14.

Study 3 Correlations Among Mathematical Outcomes

Math outcome	1.	2.	3.	4.
1. Arithmetic Fluency[a]	--			
2. Speeded Addition[b]	**.70****	--		
3. Calculation Knowledge (BMA)[b]	**.45****	**.50****	--	
4. Symbolic Number Line[a]	**.24***	**.44****	**.25***	--
5. Nonsymbolic Number Line[a]	.08	**.17***	.16†	**.47****

Note. $^{\dagger}p < .06$, $^{*}p \leq .05$, $^{**}p \leq .01$. [a]Latent factor. [b]Observed variable.

Correlations among mathematical outcomes. Correlations among mathematical outcomes are presented above in Table 3.14. Arithmetic fluency was highly correlated with speeded addition, moderately correlated with calculation knowledge, and weakly correlated with symbolic number line performance. Speeded addition was also moderately correlated with calculation knowledge and symbolic number line, but weakly correlated with nonsymbolic number line performance. Calculation knowledge was weakly correlated with both number line tasks, and the two number line tasks were moderately correlated with each other. In general, these correlations show that the outcome measures shared variance, however a substantial amount of variance remained unshared, suggesting that these outcomes tap different skills.

Correlations among PMC, executive function, and quantitative measures. The correlations among perceived math competence (PMC), executive function, and quantitative measures were the same across all sets of analyses, and thus are only described once here. PMC was correlated with executive function, symbolic comparison and ordering factors, and with the subitizing variable. The executive function factor was moderately-to-highly correlated with all quantitative measures, except for the ANS, with which it shared a smaller correlation. Generally speaking quantitative measures were moderately correlated with other quantitative measures, with some exceptions. ANS was not correlated with subitizing (consistent with H1), counting, or symbolic identification. ANS (nonsymbolic comparison large) was also more strongly correlated with nonsymbolic ordering nonsymbolic; this result was expected as this task likely taps the same underlying construct, namely, the ANS. Counting did not correlate with nonsymbolic ordering large. Symbolic identification did not correlate with nonsymbolic ordering large. The correlation between symbolic comparison and ordering factors was very high (.89).

Arithmetic Fluency SEMs

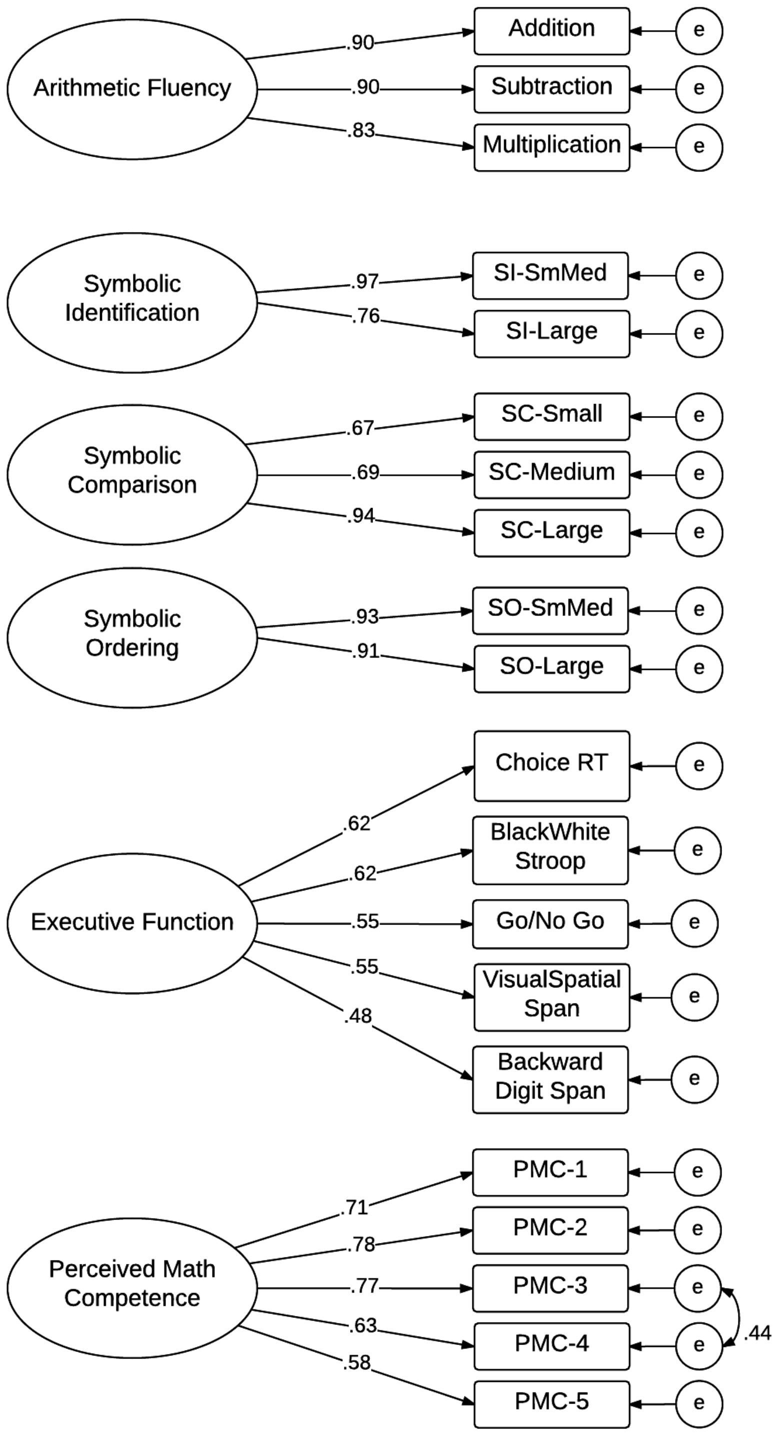
Step 3b: Correlations among predictors and arithmetic fluency. The arithmetic fluency factor was added to the existing measurement model, and model fit and factor loadings were again examined; see Figure 3.4. The factors were allowed to correlate; refer to Table 3.12 for these correlations. Model fit was good (see Table 3.13 for fit indices), and the arithmetic fluency factor loadings were ideal (all greater than .80).

PMC and executive function were moderately correlated with arithmetic fluency. All of the quantitative measures were significantly correlated with arithmetic fluency, except for ANS (nonsymbolic comparison large) and symbolic identification. The three strongest correlates of arithmetic fluency in ascending strength were subitizing, symbolic comparison, and symbolic ordering.

Step 4: Quantitative-only SEM with Arithmetic Fluency. In Step 4 the same generic "quantitative-only" SEM was conducted separately for each mathematical outcome; see Figure 3.5. Hypothesized relations among domain-specific quantitative skills were modeled in the quantitative-only SEM, and thus formally tested. To test the relative contribution of the different kinds of quantitative skills, subitizing, counting, ANS, nonsymbolic ordering large, and symbolic ordering were entered as predictors of each mathematical or number line outcome. The symbolic ordering task was selected to represent the symbolic quantitative system. The symbolic comparison and symbolic ordering factors were highly correlated, so it was unlikely that both factors would uniquely predict any outcome. ANS (nonsymbolic comparison large) and nonsymbolic ordering large were not expected to uniquely predict outcomes requiring exact calculation (i.e., arithmetic fluency, speeded addition, and calculation knowledge), but were entered as predictors to formally examine the relative contributions of the four quantitative systems because (to my knowledge) no previous study has done so.

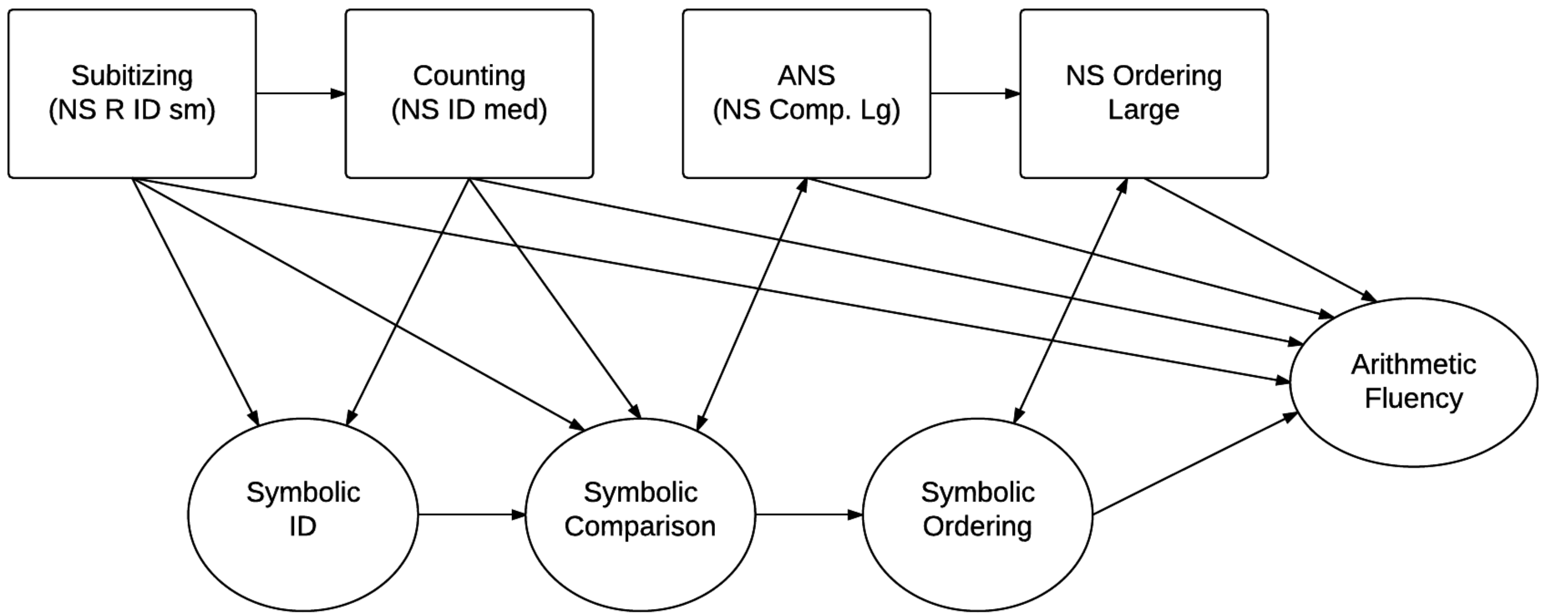
One of the advantages of SEM is that direct predictors (e.g., symbolic ordering) can be modeled to mediate relations between the outcome and other quantitative skills (e.g., symbolic comparison). Thus, by conducting SEM, I could examine whether the symbolic comparison, subitizing, and counting had significant indirect effects on the mathematical outcomes, through the constructs/measures that they predicted. The quantitative-only SEMs were conducted prior to the Full SEMs to examine the relations between the quantitative measures and mathematical outcomes prior to the addition of executive function and experiential measures. The executive function and PMC factors shared variance with the quantitative measures as well as with the math outcomes, and thus their inclusion might reduce the predictive power of the quantitative measures. Because (as far as I can tell) this is the first study to clearly operationalize and simultaneously measure subitizing, counting, ANS, and symbolic quantitative skills among adults, it seemed worthwhile to examine them in relation to each other and in relation to mathematical outcomes prior to complicating the model, and perhaps attenuating some of the observed effects.

Figure 3.4. Measurement model with arithmetic fluency as the outcome.



Note: Standardized solution shown. Factors were set to be correlated; correlations are not shown in the figure, but rather presented in Table 3.13. Factor loadings were all significant at $p < .001$.

Figure 3.5. The generic "Quantitative-Only" SEM: Quantitative skills only as predictors of the mathematical outcome.

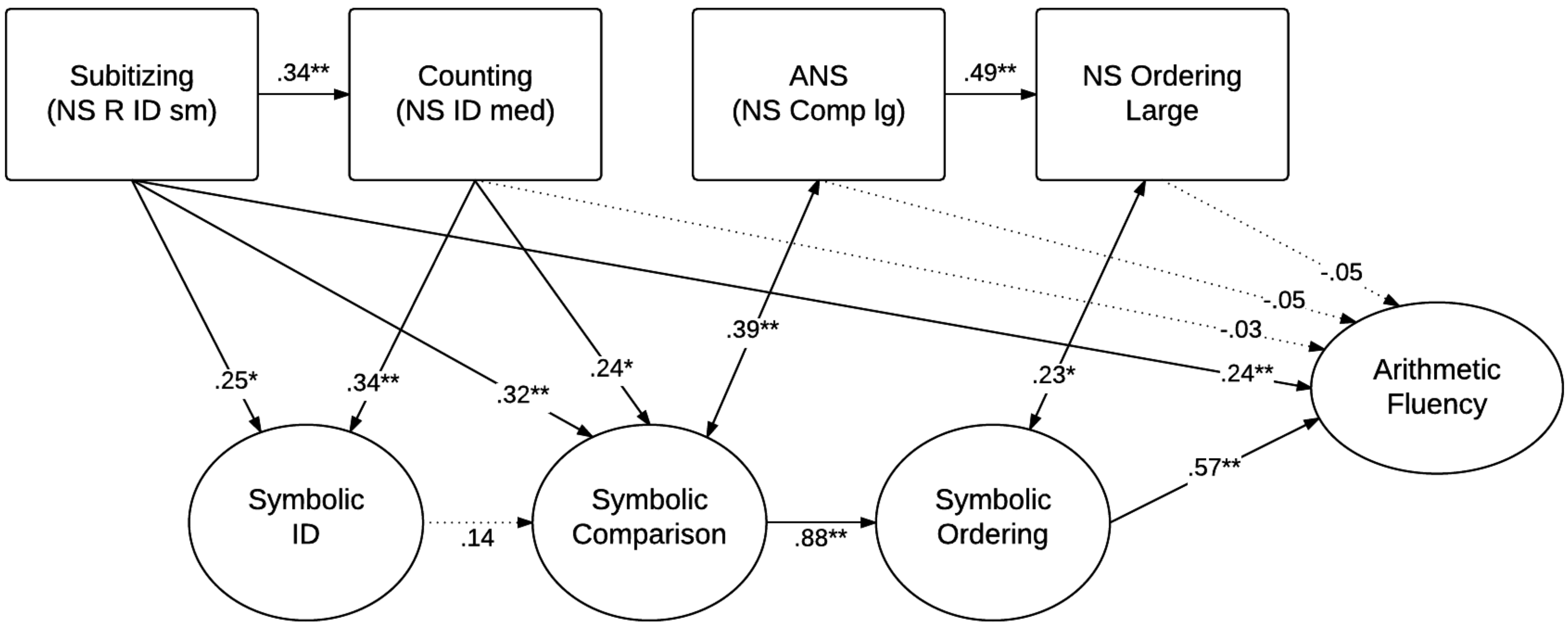


Note: Arithmetic fluency was substituted for other mathematical outcomes at Step 4 in each set of analyses.
NS = Nonsymbolic. R = Rapid. ID = Identification. Comp = Comparison. sm = small. med = medium. lg = large.

Figure 3.6 shows the standardized solution for the quantitative-only SEM with arithmetic fluency as the outcome. The relations described here were constant across all of the quantitative-only SEMs conducted at Step 4. As hypothesized, subitizing was a significant predictor of counting (H3), and symbolic identification and comparison (H4). Counting also predicted symbolic identification and comparison (H7). The hypothesis that simpler symbolic quantitative skills would predict more complicated ones was partially supported (H9). Symbolic identification did not predict symbolic comparison, but symbolic comparison was a strong predictor of symbolic ordering. ANS (nonsymbolic comparison large) was also predictive of nonsymbolic ordering large. Symbolic and nonsymbolic (ANS) comparison tasks were significantly correlated, as were symbolic and nonsymbolic ordering. The variances accounted for in each endogenous variable (for all SEMs) are shown in Table 3.15.

The arithmetic fluency quantitative-only SEM was a good fit to the data (refer to Table 3.13 for all final model fit indices); see Figure 3.6 for estimated regression coefficients. Direct and indirect effects for all mathematical outcomes are shown in Table 3.16 (in separate columns for quantitative-only and full SEMs). The significant quantitative predictors of arithmetic fluency were symbolic ordering and subitizing. As expected, symbolic ordering was the strongest quantitative predictor of arithmetic fluency (H11). Inconsistent with the hypothesis that nonsymbolic quantitative skill would not predict mathematical outcomes beyond symbolic quantitative skill (H10), subitizing was also a unique predictor of arithmetic fluency. Note that this finding is consistent with Study 2 path model results. Subitizing, counting, and symbolic comparison also had significant indirect effects on arithmetic fluency (effects displayed in Table 3.16).

Figure 3.6. Arithmetic Fluency SEM: Quantitative predictors only.



Note. [*]$p < .05$, [**]$p \leq .001$. $N = 153$. Values are standardized; measurement model is not shown.
NS = Nonsymbolic. R = Rapid. ID = Identification. Comp = Comparison. sm = small. med = medium. lg = large.
Arithmetic fluency $R^2 = .45$.

Table 3.15.

Study 3 SEM results: Total Variance Accounted For (R^2s)

Endogenous Variables	Model: Quantitative Only	Model: Full SEM
Mathematical outcomes		
Arithmetic Fluency	**.45****	**.63****
Speeded Addition	**.50****	**.54****
Calculation Knowledge	**.17****	**.30****
Symbolic Number Line	**.29****	**.32****
Nonsymbolic Number Line	**.24****	**.29****
Nonsymbolic Quantitative		
Subitizing	**--**	**.42****
Counting	**.11***	**.24****
Ordering large	**.24****	**.32****
Symbolic Quantitative		
Identification	**.23****	**.35****
Comparison	**.29****	**.45****
Ordering	**.77***	**.78****

Note: $^{*}p \leq .05$, $^{**}p \leq .001$. Values are R^2s.
Quantitative Only = SEMs with quantitative predictors and the mathematical outcome only. The variance accounted for in the symbolic and nonsymbolic quantitative measures was consistent across models/outcomes.
ANS was the only exogenous quantitative measure. Subitizing was exogenous in Quantitative-Only model.

Table 3.16. *Summary of Direct and Indirect Effects Across Mathematical Outcomes.*

	Arithmetic Fluency		Speeded Addition		Calc.Knowledge		Symb Number Line		NSymb Number Line	
	Q only	Full SEM	Q only	Full SEM	Q only	Full SEM	Q only	Full SEM	Q only	Full SEM
Subitizing Total	$\mathbf{.45}^{**}$	$\mathbf{.24}^{**}$	$\mathbf{.36}^{**}$	.07	.11	-.05	.11	-.04	.03	-.02
Direct	$\mathbf{.24}^{**}$	$\mathbf{.21}^{**}$	.06	.03	-.03	-.06	-.08	-.07	-.04	-.03
Indirect	$\mathbf{.21}^{**}$	.03	$\mathbf{.30}^{**}$	.04	$\mathbf{.14}^{*}$	.01	$\mathbf{.18}^{**}$	.03	.07	.01
Counting Total	.11	.07	$\mathbf{.16}^{*}$	.05	.03	.01	.14	.08	$\mathbf{.16}^{*}$	$\mathbf{.16}^{*}$
Direct	-.03	.01	-.05	-.03	-.08	-.04	.03	.02	$\mathbf{.15}^{\dagger}$	$\mathbf{.14}^{\dagger}$
Indirect	$\mathbf{.14}^{**}$	.05	$\mathbf{.20}^{**}$	.08	$\mathbf{.11}^{*}$	.03	$\mathbf{.11*}$	.05	.01	.01
ANS Total	-.07	-.09	$\mathbf{-.13}^{*}$	$\mathbf{-.13*}$	.00	.02	$\mathbf{.19*}$	$\mathbf{.18*}$	$\mathbf{.41}^{**}$	$\mathbf{.37}^{**}$
Direct	-.05	-.08	-.12	-.12	-.03	-.01	.12	.11	$\mathbf{.29}^{**}$	$\mathbf{.27}^{**}$
Indirect	-.02	-.01	-.01	.01	.03	.03	.08	.07	$\mathbf{.12}^{*}$	$\mathbf{.10}^{*}$
NSymb Order Lg Total	-.05	-.04	-.03	-.02	.08	.08	$.16^{\dagger}$	$.16^{\dagger}$	$\mathbf{.24}^{*}$	$\mathbf{.23}^{*}$
Direct	-.05	-.04	-.03	-.02	.08	.08	$.16^{\dagger}$	$.16^{\dagger}$	$\mathbf{.24}^{*}$	$\mathbf{.23}^{*}$
Indirect	--	--	--	--	--	--	--	--	--	--
Symbolic ID Total	.07	-.01	.10	-.01	.10	.00	.05	-.01	.00	.00
Direct	--	--	--	--	--	--	--	--	--	--
Indirect	.07	-.01	.10	-.01	.10	.00	.05	-.01	.00	.00
Symbolic Comp. Total	$\mathbf{.50}^{**}$	$\mathbf{.38}^{**}$	$\mathbf{.70}^{**}$	$\mathbf{.61}^{**}$	$\mathbf{.38}^{**}$	$\mathbf{.20}^{*}$	$\mathbf{.38}^{**}$	$\mathbf{.40}^{**}$	.03	.11
Direct	--	--	$\mathbf{.37}^{**}$	$\mathbf{.44}^{**}$	--	--	--	--	--	--
Indirect	$\mathbf{.50}^{**}$	$\mathbf{.38}^{**}$	$\mathbf{.33}^{**}$	.17	$\mathbf{.38}^{**}$	$\mathbf{.20}^{*}$	$\mathbf{.38}^{**}$	$\mathbf{.40}^{**}$	.03	.11
Symbolic Ordering Total	$\mathbf{.57}^{**}$	$\mathbf{.43}^{**}$	$\mathbf{.37}^{**}$	.19	$\mathbf{.43}^{**}$	$\mathbf{.22}^{**}$	$\mathbf{.43}^{**}$	$\mathbf{.45}^{**}$	.04	.12
Direct	$\mathbf{.57}^{**}$	$\mathbf{.43}^{**}$	$\mathbf{.37}^{**}$	.19	$\mathbf{.43}^{**}$	$\mathbf{.22}^{**}$	$\mathbf{.43}^{**}$	$\mathbf{.45}^{**}$	.04	.12
Indirect	--	--	--	--	--	--	--	--	--	--
Executive Function Total	--	$\mathbf{.38}^{**}$	--	$\mathbf{.40}^{**}$	--	.08	--	$\mathbf{.27}^{**}$	--	$\mathbf{.19}^{**}$
Direct	--	--	--	--	--	--	--	--	--	--
Indirect	--	$\mathbf{.38}^{**}$	--	$\mathbf{.40}^{**}$	--	.08	--	$\mathbf{.27}^{**}$	--	$\mathbf{.19}^{**}$

Note: $^{\dagger}p < .07$; $^{*}p \leq .05$, $^{**}p \leq .001$. Q only = SEM with quantitative predictors and outcome only. Values are standardized coefficients.

Step 5: Full SEM predicting arithmetic fluency. The full SEM model was first established with arithmetic fluency, and then specified in the same way for each subsequent mathematical outcome. PMC, age, and education location were added as predictors of the mathematical outcome, and executive function was entered as a predictor of subitizing, counting, and symbolic identification. PMC was allowed to correlate with executive function, symbolic comparison, and symbolic ordering, as it was moderately correlated with these factors in the measurement model. Executive function was also correlated with ANS default by MPlus default (exogenous constructs are set to be correlated to help to stabilize the model)[7]. The initial model fit was less than ideal, $\chi^2(276) = 491.6$, $p < .001$, CFI = .889, TLI = .871, SRMR = .093, RMSEA = .071 (LCI .061, UCI = .081). Modification indices suggested adding a correlation between PMC and age, and model fit improved slightly, $\chi^2(277) = 479.1$, $p < .001$, CFI = .895, TLI = .878, SRMR = .090, RMSEA = .069 (LCI = .059, UCI = .079).

Executive function was correlated with the other quantitative measures and modification indices suggested adding it as a predictor of symbolic comparison and nonsymbolic ordering large. The model fit improved but remained less than ideal; see final model fit information in Table 3.13. No other modification indices were theoretically sound, nor were they associated with large changes in model fit, thus the model was deemed suitable and results are presented in Figure 3.7. Lack of model fit can be partially attributed to the 99 remaining free parameters and the fairly large model being tested. Despite the poor fit, most specified pathways were significant, and factor loadings were similar to those in the measurement model. Furthermore, SEM analyses

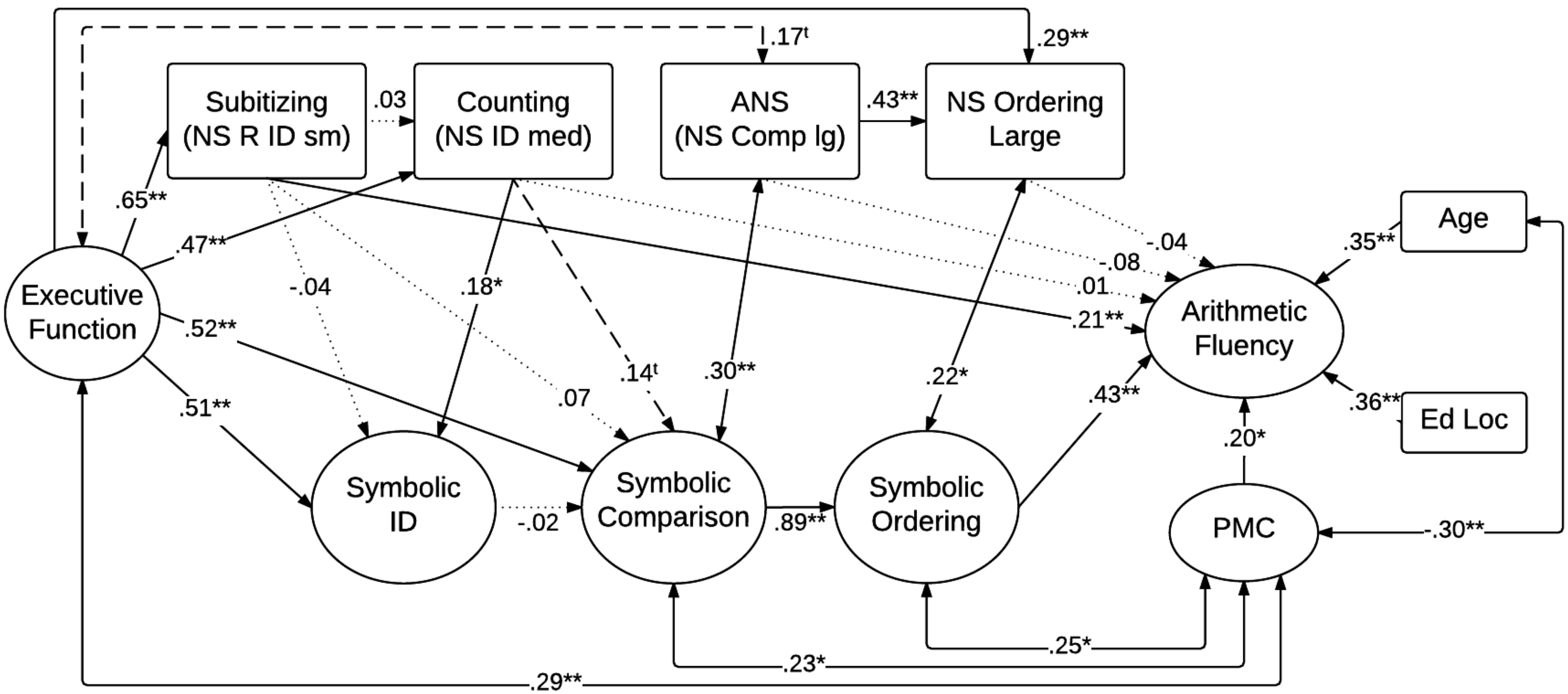
[7] Other default correlations were set to zero as they did not fit with theory or with the observed correlations (e.g., by default, ANS would correlate with PMC; this was set to 0, as observed correlation was .05). The presentation of full SEMs is transparent; correlations are shown in figures if present.

allowed for simultaneous analysis of direct and indirect predictors of arithmetic fluency.

Executive function was a strong predictor of the quantitative measures in all of the Full SEMS, supporting the hypothesis that domain-general cognitive skills contribute to the development of domain specific quantitative skills (H12). Once executive function was added to the model, subitizing did not predict the quantitative skills that it had predicted in the Quantitative Only SEMS (i.e., counting, symbolic identification, and symbolic comparison). Counting was also weakened as a predictor of symbolic identification and comparison. It seems that the variance subitizing shared with the other quantitative factors overlapped with the variance shared by subitizing and executive function; only executive function predicted these quantitative measures once included.

Despite no longer being a significant predictor of other quantitative skills, subitizing remained a significant predictor of arithmetic fluency. Symbolic ordering was the strongest quantitative predictor, as hypothesized (H11). PMC, age, and education location were also significant predictors (supporting H13). Executive function had a significant indirect effect on arithmetic fluency ($\beta = .38$, $p < .001$; H12), primarily through the path from symbolic comparison to symbolic ordering to arithmetic fluency. Very little of executive function's indirect effect was associated with the paths going through subitizing—which means that the variance shared among subitizing and executive function was separate from the variance shared between subitizing and arithmetic fluency. The indirect effect of symbolic comparison on arithmetic fluency was significant through symbolic ordering ($\beta = .38$, p < .001). See Table 3.15 for the variance accounted for in each mathematical outcomes and each endogenous quantitative variables. The variance accounted for in arithmetic fluency was $R^2 = .63$.

Figure 3.7. Arithmetic Fluency Full SEM.



Note. $^{t}p < .09$, $^{*}p < .05$, $^{**}p \leq .001$. $N = 153$. Values are standardized; measurement model is not shown.
NS = Nonsymbolic. R = Rapid. ID = Identification. Comp = Comparison. sm = small. med = medium. lg = large.
Ed loc = Education location. PMC = Perceived math competence. Arithmetic fluency $R^2 = .63$.

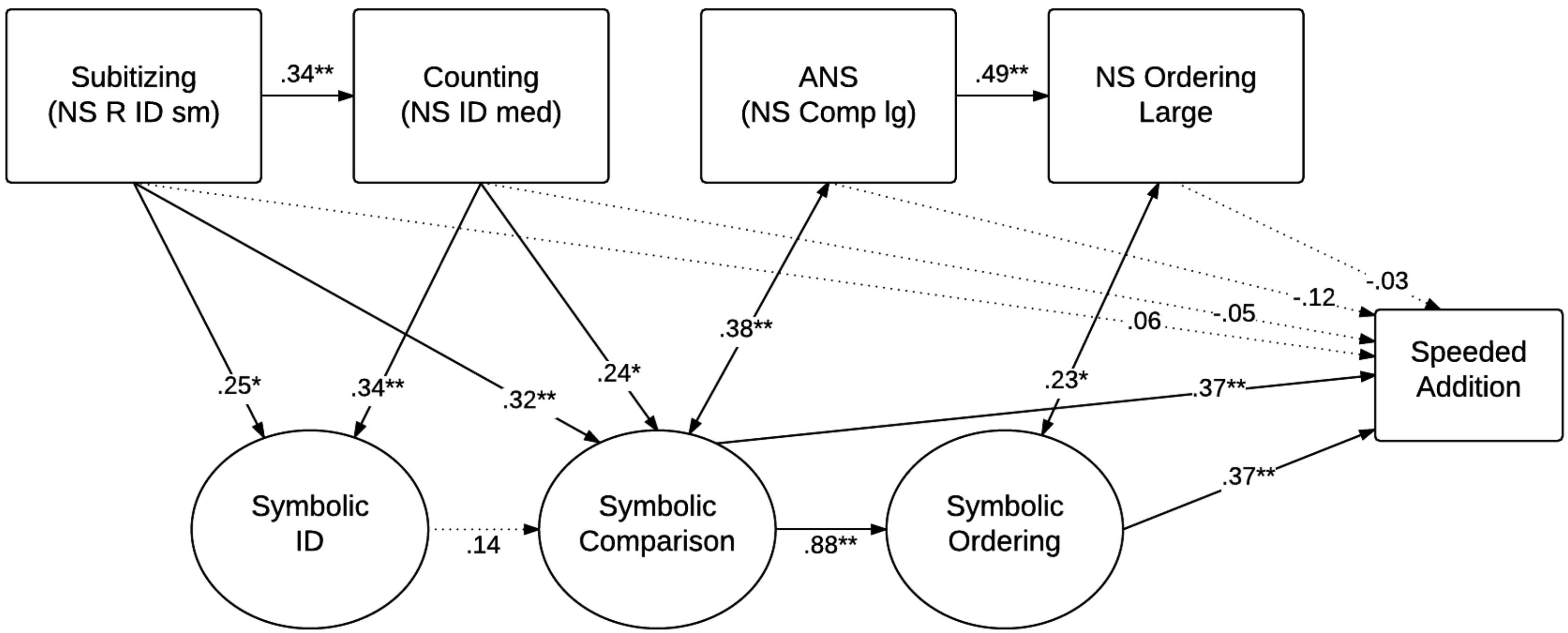
Speeded Addition SEMs

Step 3b: Correlations among predictors and speeded addition. Speeded addition was correlated with all of the predictor variables (Table 3.12). Correlations were moderate with PMC, executive function, subitizing, and counting, and smaller with nonsymbolic ordering large, and symbolic identification. The strongest correlates with speeded addition were the symbolic comparison and ordering factors.

Step 4: Quantitative-only SEM with speeded addition. The speeded addition quantitative-only SEM was a good fit to the data (see Table 3.13) and accounted for 50% of the variance in speeded addition (R^2s in Table 3.16); see Figure 3.8 for the standardized solution. Significant quantitative predictors of speeded addition were symbolic comparison and symbolic ordering. Nonsymbolic quantitative skills were not uniquely predictive of speeded addition (consistent with Hypothesis 10). Subitizing and counting were indirectly related to speeded addition; see Table 3.16 for indirect effects.

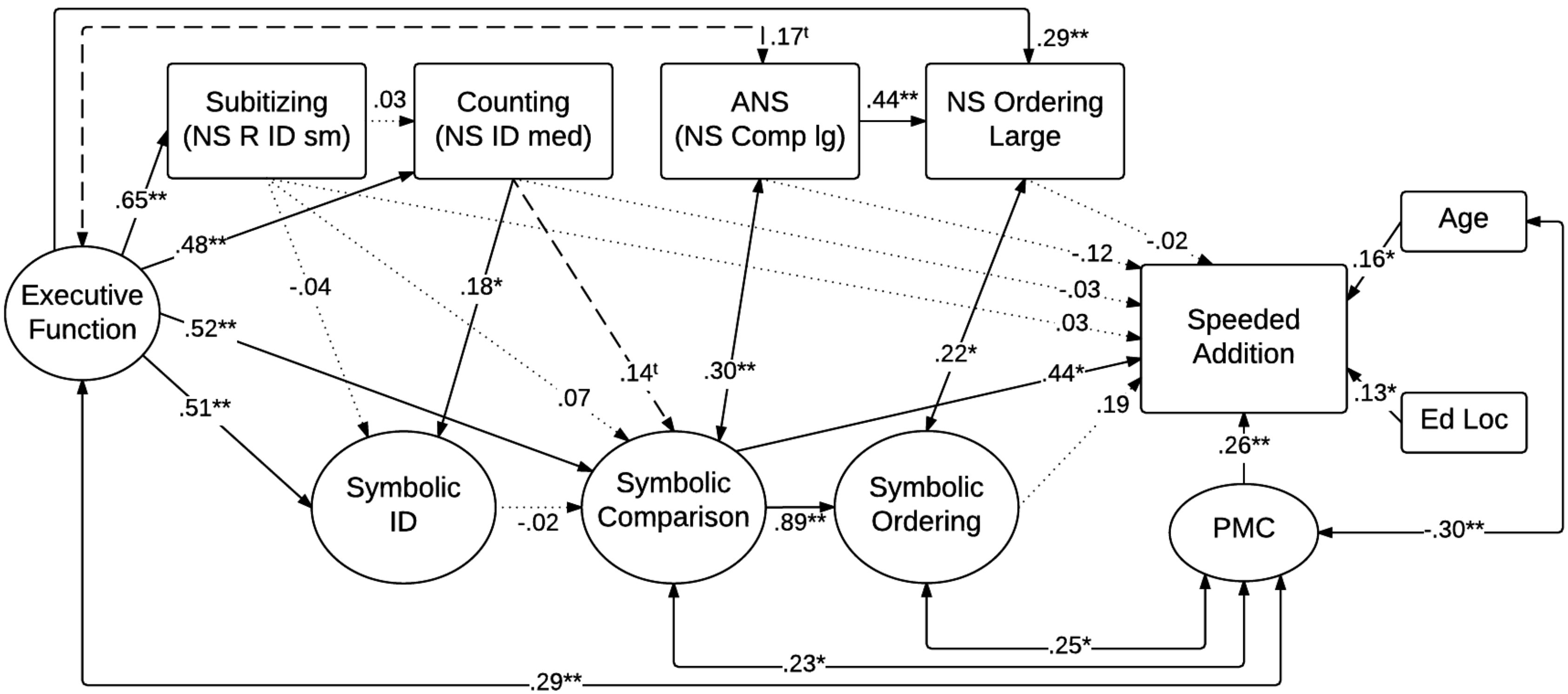
Step 5: Full SEM predicting speeded addition. See Figure 3.9 for the full SEM model predicting speeded addition. The variance accounted for in speeded addition was $R^2 = .54$. Inconsistent with the Hypothesis 11, symbolic comparison—not symbolic ordering—was the best quantitative predictor of speeded addition. Of the two, symbolic comparison was the only significant direct predictor, and its total effect (including the indirect effect) was .61. PMC, age, and education location were also significant predictors (supporting H13), although the effects of education location and age were much smaller relative to those observed in the arithmetic fluency full SEM. As expected, executive function had a significant indirect effect on speeded addition (H12).

Figure 3.8. Speeded Addition SEM: Quantitative predictors only.



Note. [*]$p < .05$, [**]$p \leq .001$. $N = 149$. Values are standardized; measurement model is not shown.
NS = Nonsymbolic. R = Rapid. ID = Identification. Comp = Comparison. sm = small. med = medium. lg = large.
Speeded Addition $R^2 = .50$.

Figure 3.9. Speeded Addition Full SEM.



Note. $^{t}p < .09$, $^{*}p < .05$, $^{**}p \leq .001$. $N = 149$. Values are standardized; measurement model is not shown.
NS = Nonsymbolic. R = Rapid. ID = Identification. Comp = Comparison. sm = small. med = medium. lg = large.
Ed loc = Education location. PMC = Perceived math competence. Speeded Addition $R^2 = .54$

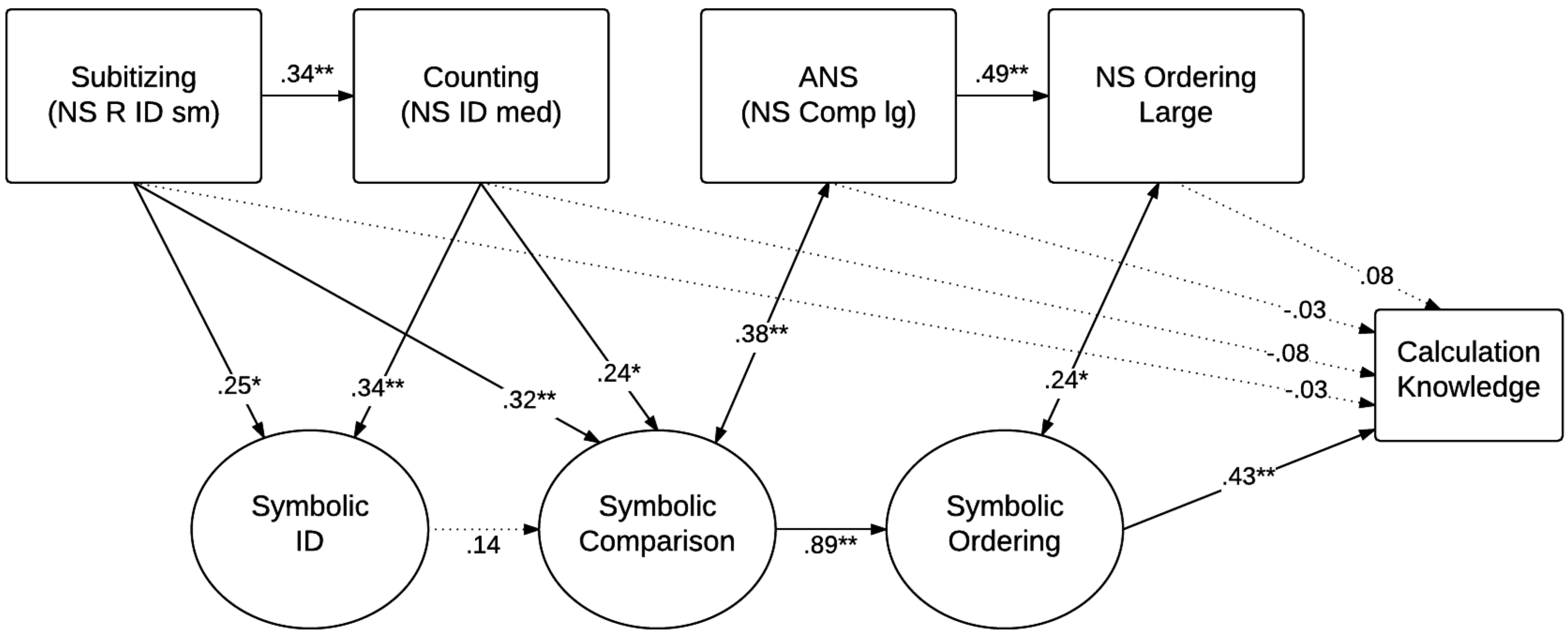
Calculation Knowledge SEMs

Step 3b: Correlations among predictors and calculation knowledge. Calculation knowledge (as measured by the Brief Math Assessment) was moderately correlated with the PMC, symbolic comparison, and symbolic ordering factors, and shared small correlations with executive function and nonsymbolic ordering large (correlations in Table 3.12). Unlike arithmetic fluency and speeded addition, calculation knowledge was not significantly correlated with subitizing or counting.

Step 4: Quantitative-only SEM with calculation knowledge. The calculation knowledge quantitative-only SEM was a good fit to the data, however it only accounted for 17% of the variance in calculation knowledge—much less than variance accounted for in arithmetic fluency and speeded addition at Step 4 (see Table 3.15). See Figure 3.10 for estimated regression coefficients and Table 3.16 for direct and indirect effects. Consistent with hypotheses, nonsymbolic quantitative skills were not predictive of calculation knowledge (H10), and symbolic ordering was the only (and therefore best) significant quantitative predictor (H11). Subitizing and counting also demonstrated small but significant indirect effects on calculation knowledge in Step 4.

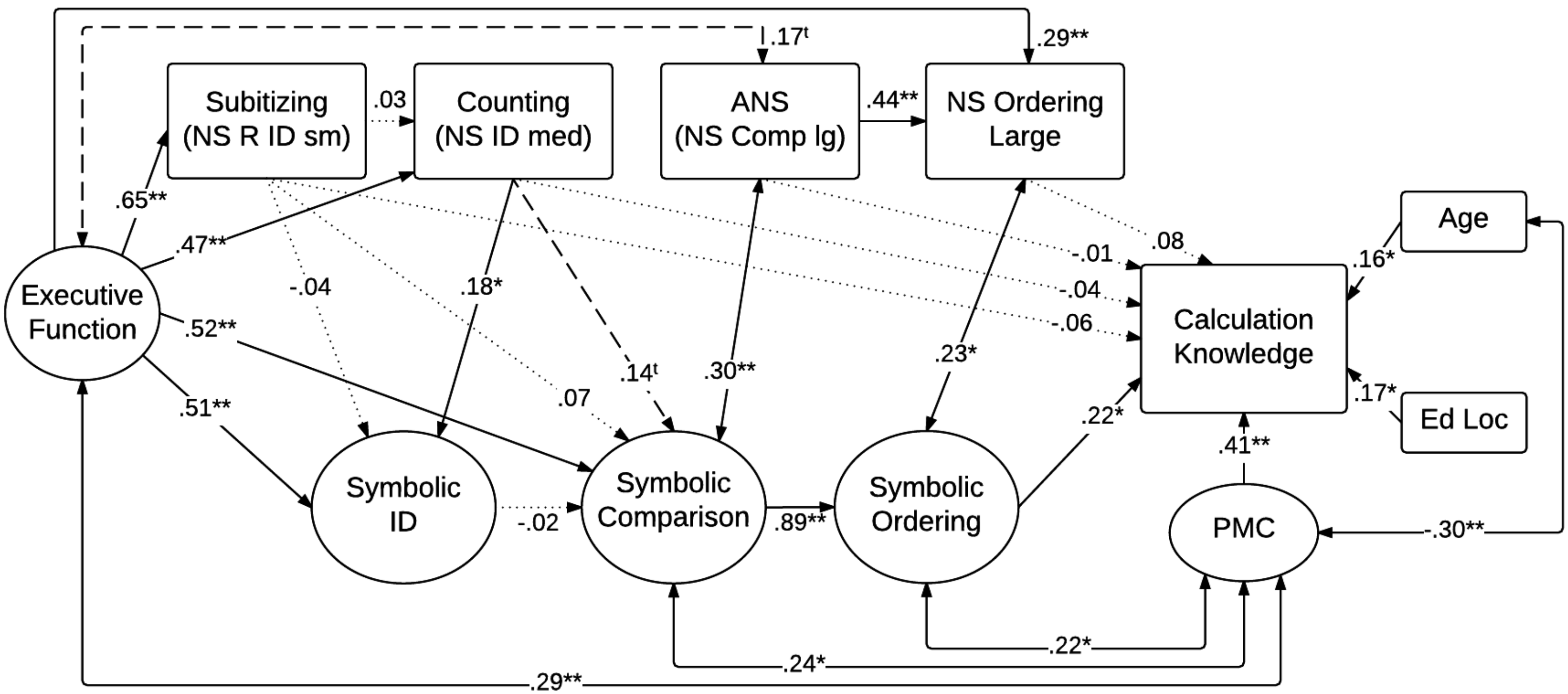
Step 5: Full SEM predicting calculation knowledge. See Figure 3.11 for the full SEM model predicting calculation knowledge. Symbolic ordering remained a significant predictor. PMC, Age, and education location were also significant (supporting H13). The effects sizes for age and education were consistent with those observed in the speeded addition full SEM (and smaller than those observed in the arithmetic fluency full SEM). Compared to arithmetic fluency and speeded addition models, PMC was a stronger predictor of calculation knowledge. The predictive strength of symbolic ordering

Figure 3.10. Calculation Knowledge SEM: Quantitative predictors only.



Note. $^{*}p < .05$, $^{**}p \leq .001$. $N = 153$. Values are standardized; measurement model is not shown.
NS = Nonsymbolic. R = Rapid. ID = Identification. Comp = Comparison. sm = small. med = medium. lg = large.
Calculation Knowledge $R^2 = .17$.

Figure 3.11. Calculation Knowledge Full SEM.



Note. $^{t}p < .09$, $^{*}p < .05$, $^{**}p \leq .001$. $N = 153$. Values are standardized; measurement model is not shown. NS = Nonsymbolic. R = Rapid. ID = Identification. Comp = Comparison. sm = small. med = medium. lg = large. Ed loc = Education location. PMC = Perceived math competence. Calculation Knowledge $R^2 = .30$.

was also smaller in the calculation knowledge model relative to the arithmetic fluency model. Executive function was not indirectly related to calculation knowledge (contrary to H12). Not surprisingly, the variance accounted for in calculation knowledge was also less (R^2 = .30) compared to arithmetic fluency and speeded addition.

Symbolic Number Line SEMs

Step 3b: Correlations among predictors and symbolic number line. A symbolic number line factor with three indicators was added to the measurement model; the current measurement model is shown in Figure 3.12. Factors were allowed to correlate; correlations are shown in Table 3.12 rather than in the figure. This measurement model was a good fit to the data (see Table 3.13) and symbolic number line factor loadings were adequate. Symbolic number line performance was correlated with all of the quantitative measures except for symbolic identification; its correlations with symbolic comparison and ordering were strongest. Also of note, symbolic number line was correlated with ANS whereas the previous mathematical outcomes were not; it also shared a stronger correlation with nonsymbolic ordering than did the exact-calculation measures (approximate r = .20 versus r = .36). Symbolic number line performance was also correlated with PMC and executive function. These correlations were consistent with the hypothesis that ANS would not uniquely predict (exact) arithmetic-based outcomes—but would predict numerical outcomes scored in terms of approximate correctness.

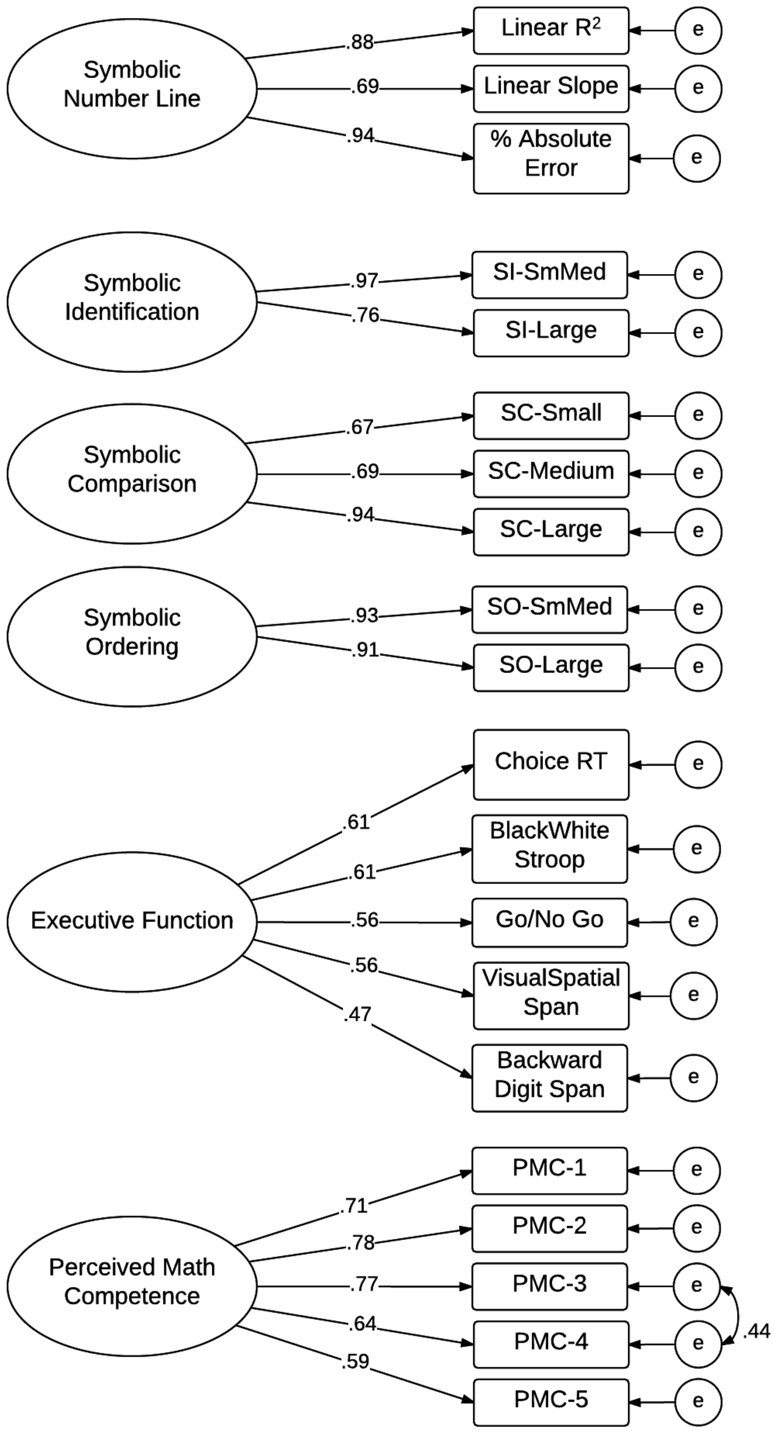
Step 4: Quantitative-only SEM with symbolic number line. The quantitative-only SEM for symbolic number line is shown in Figure 3.13, which was a good fit to the data. This model accounted for 29% of the variance in symbolic number line performance. Symbolic ordering was a significant unique predictor of symbolic number line performance; nonsymbolic ordering was a marginally significant predictor as well. The indirect effects of subitizing, counting, symbolic comparison were significant, and the ANS total effect (direct plus indirect) was significant as well. Thus, there was some support for the Hypothesis 6 that ANS would predict the number line tasks, which do not require exact calculation.

Step 5: Full SEM predicting symbolic number line. See Figure 3.14 for the full SEM model predicting symbolic number line performance. In contrast to the previous full SEMs, PMC, education location, and age did not predict performance on this outcome. Symbolic ordering and nonsymbolic ordering (marginal) remained significant predictors. Symbolic comparison had a significant indirect effect, as did executive function (H12). The total effect of ANS (direct plus indirect) was also significant. The variance accounted for in symbolic number line performance was R^2 =.32 (see Table 3.16).

Nonsymbolic Number Line SEMs

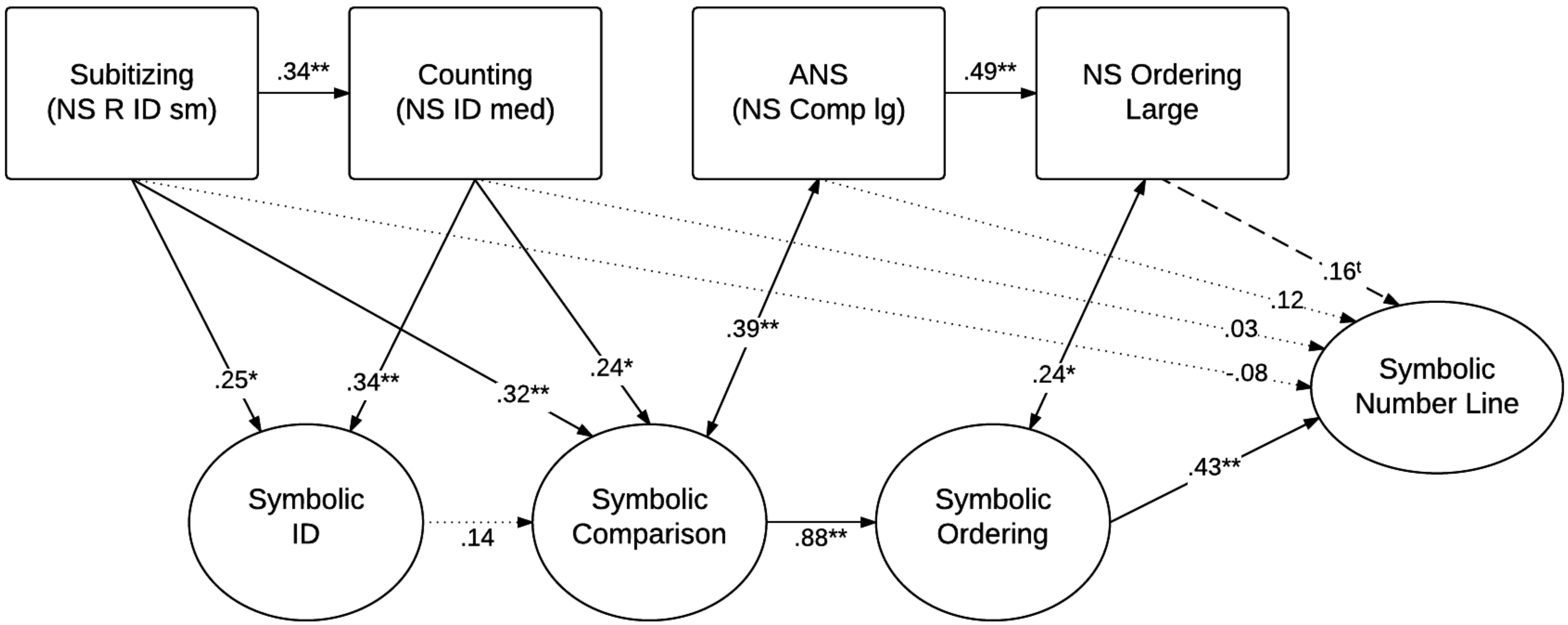
Step 3b: Correlations among predictors and nonsymbolic number line. A nonsymbolic number line factor with three indicators was added to the measurement model, which is shown in Figure 3.15. Factors were allowed to correlate; correlations are shown in Table 3.12 rather than in the figure. This measurement model was a good fit to the data (see Table 3.13) and nonsymbolic number line factor loadings were ideal (all above .85, two above .90). Nonsymbolic number line performance was moderately correlated with executive function, ANS and nonsymbolic ordering large (these were a little stronger than for symbolic number line). Nonsymbolic number line performance also shared small correlations with counting, symbolic comparison, and symbolic ordering.

Figure 3.12. Measurement model with symbolic number line as the outcome.



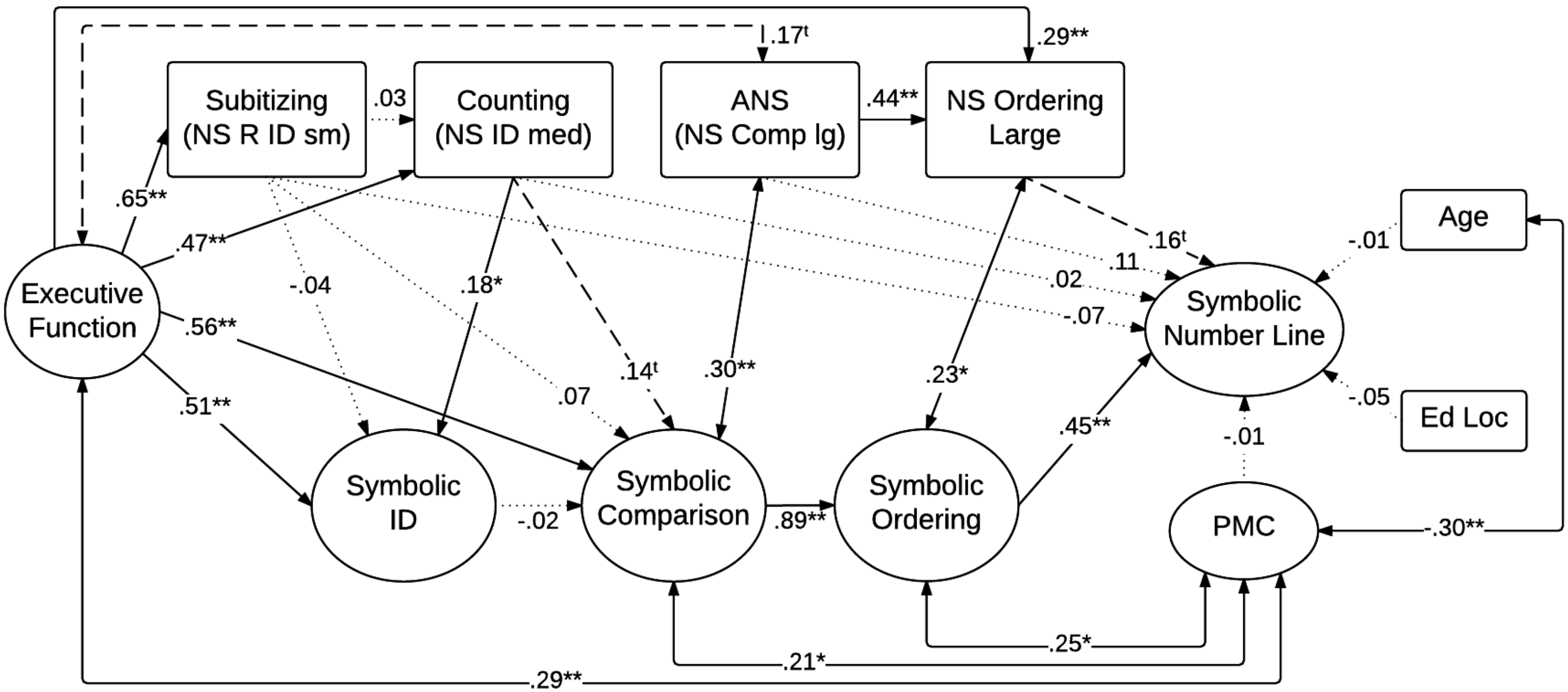
Note: The standardized solution is shown. Factors were allowed to correlate. Correlations are presented in Table 3.12 rather than shown here. Factor loadings were all significant at $p < .001$.

Figure 3.13. Symbolic Number Line SEM: Quantitative predictors only.



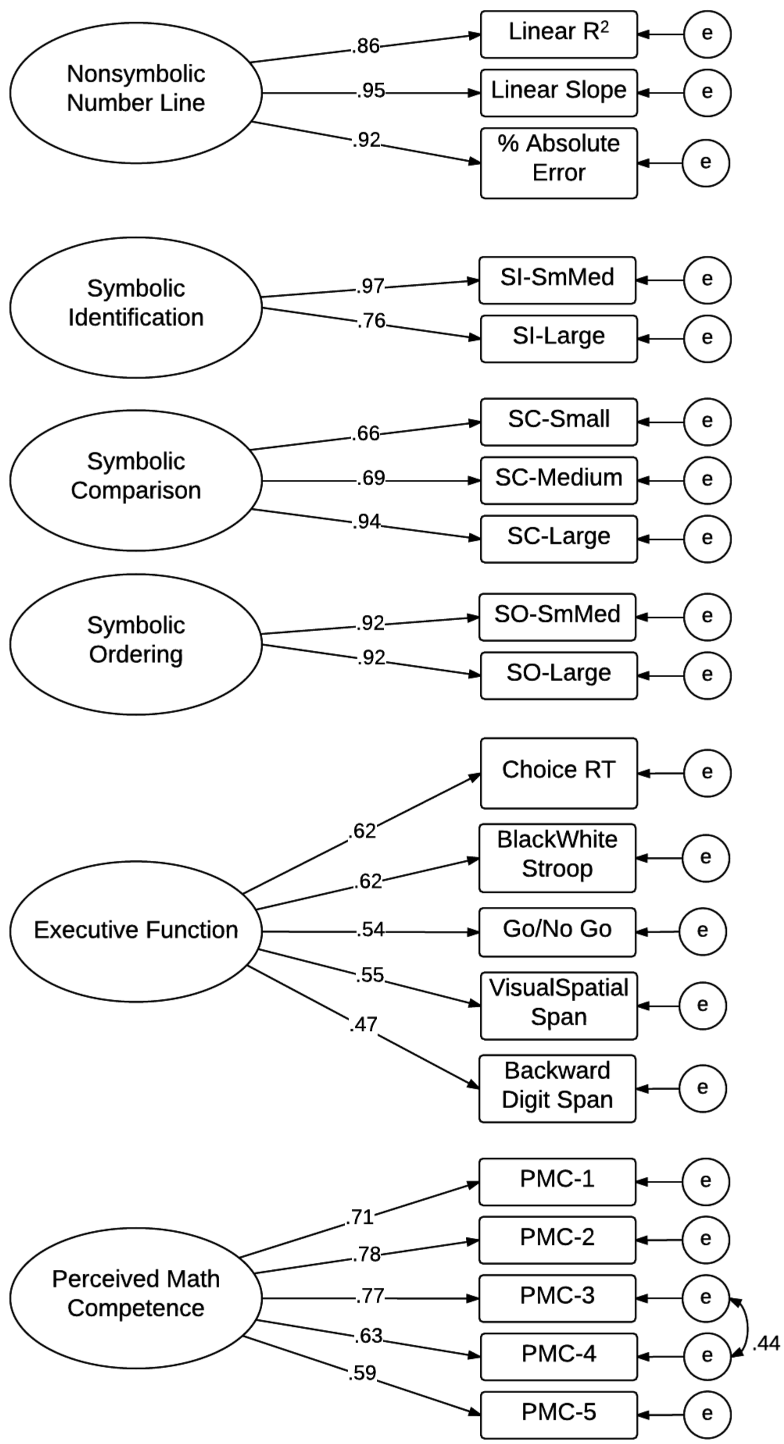
Note. $^{*}p < .05$, $^{**}p \leq .001$. $N = 150$. Values are standardized; measurement model is not shown.
NS = Nonsymbolic. R = Rapid. ID = Identification. Comp = Comparison. sm = small. med = medium. lg = large.
Symbolic number line $R^2 = .29$.

Figure 3.14. Symbolic Number Line Full SEM.



Note. ${}^{t}p < .09$, ${}^{*}p < .05$, ${}^{**}p \leq .001$. $N = 150$. Values are standardized; measurement model is not shown.
NS = Nonsymbolic. R = Rapid. ID = Identification. Comp = Comparison. sm = small. med = medium. lg = large.
Ed loc = Education location. PMC = Perceived math competence. Symbolic number line $R^2 = .32$.

Figure 3.15. Measurement model with nonsymbolic number line as the outcome.



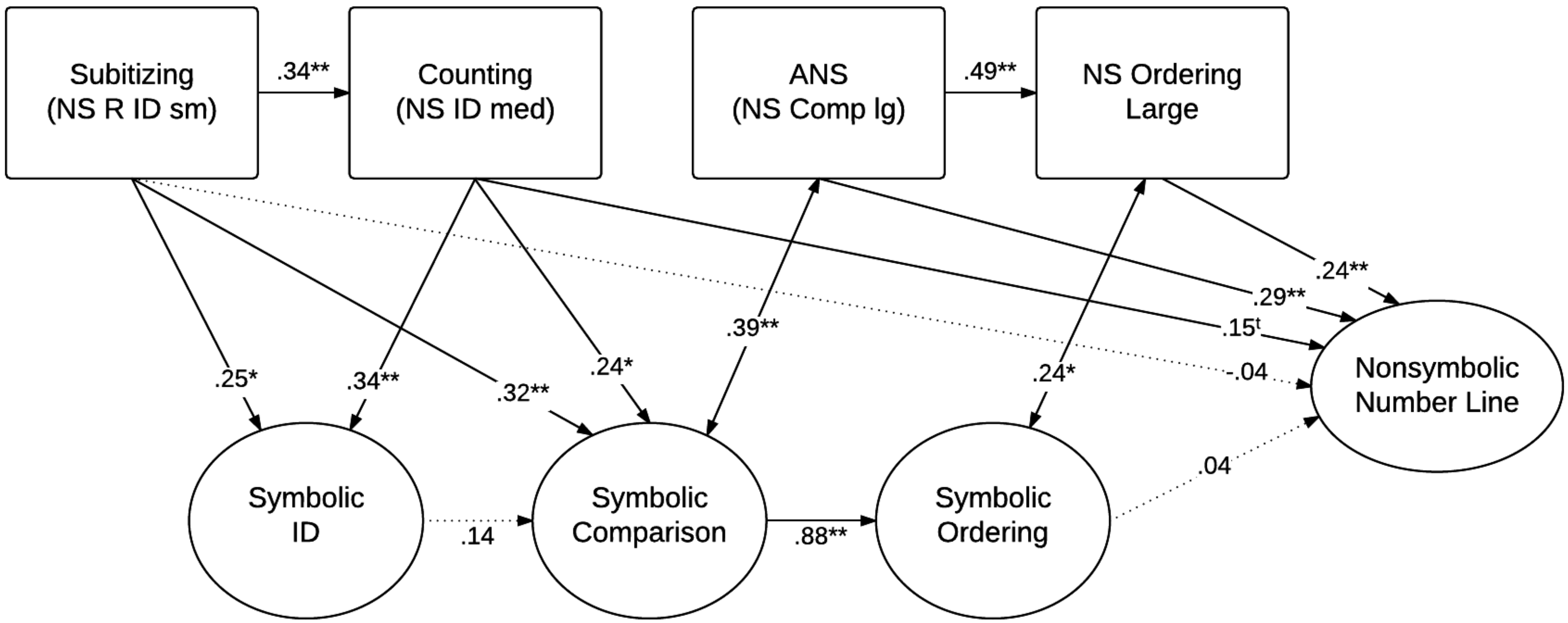
Note: Standardized solutions shown. Factors were allowed to correlate. Correlations are presented in Table 3.12 rather than shown here. Factor loadings all significant at $p < .001$.

Step 4: Quantitative-only SEM with nonsymbolic number line. The quantitative-only SEM for nonsymbolic number line is shown in Figure 3.16 and was a good fit to the data. This model only accounted for 24% of the variance in nonsymbolic number line performance (similar to symbolic number line and calculation knowledge). ANS and nonsymbolic ordering large were significant unique predictors of symbolic number line performance (supporting H6), and counting was marginally significant. ANS also had a significant indirect effect (total effect = .41).

Step 5: Full SEM predicting nonsymbolic number line. See Figure 3.17 for the Full SEM model predicting nonsymbolic number line performance. Consistent with the simple correlations, age and education location were not predictive of nonsymbolic number line performance. PMC was a marginally significant, but negative, predictor. ANS and nonsymbolic ordering large remained significant predictors, and counting a marginally significant predictor, of nonsymbolic number line performance. Executive function also had a significant indirect effect (supporting H12). The variance accounted for in nonsymbolic number line performance was R^2 = .29 (R^2s shown in Table 3.15).

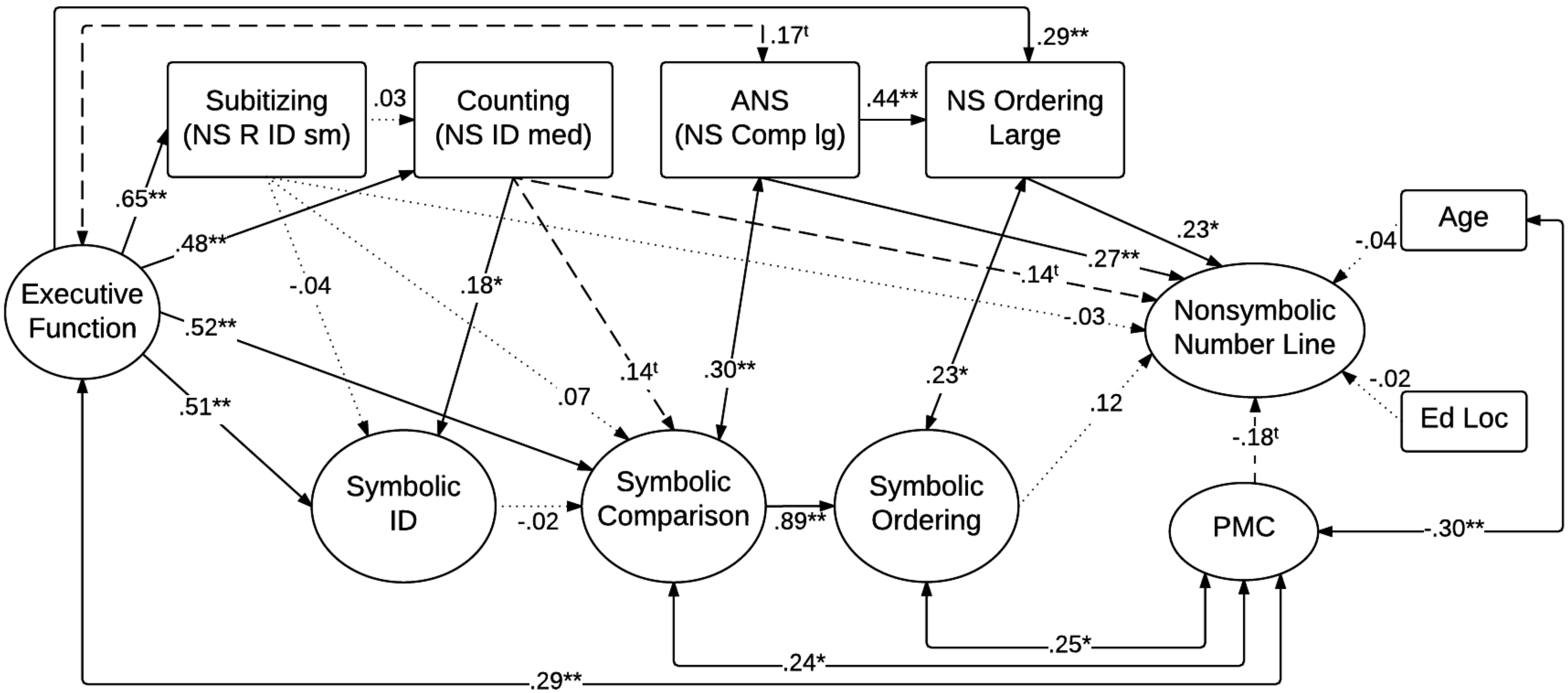
.

Figure 3.16. Nonsymbolic Number Line SEM: Quantitative predictors only.



Note. $^{t}p < .09$, $^{*}p < .05$, $^{**}p \leq .001$. Values are standardized; measurement model is not shown.
NS = Nonsymbolic. R = Rapid. ID = Identification. Comp = Comparison. sm = small. med = medium. lg = large.
Nonymbolic number line $R^2 = .24$.

Figure 3.17. Nonsymbolic number line Full SEM.



Note. $^{t}p < .09$, $^{*}p < .05$, $^{**}p \leq .001$. $N = 153$. Values are standardized; measurement model is not shown.
NS = Nonsymbolic. R = Rapid. ID = Identification. Comp = Comparison. sm = small. med = medium. lg = large.
Ed loc = Education location. PMC = Perceived math competence. Nonsymbolic number line $R^2 = .29$.

Discussion

Study 3 included a comprehensive set of quantitative tasks in order to systematically examine theory-based predictions about the relations between quantitative processes and mathematical outcomes in adults. The quantitative tasks selected for inclusion in this study varied on three dimensions: format (symbolic [digits] or nonsymbolic [dots]); range of the quantities (small [1 to 3], medium [5 to 9], and large [10 to 100]); and task (identification, comparison, and ordering). One of the main goals was to separately measure the subitizing, counting, and approximate number systems with nonsymbolic quantitative tasks; analogous symbolic quantitative skills were also measured. I was thus permitted to examine subitizing, counting, ANS, and symbolic quantitative performance: a) in relation to other quantitative skills, b) in relation to multiple mathematical outcomes, c) and in relation to domain-general cognitive skills. Here I discuss the findings of Study 3 in three parts. First, potential methodological issues are discussed. Second, the differences and similarities between the Quantitative-Only and Full structural equation models are discussed. Third, I summarize the overall findings and indicate whether they supported my specific hypotheses. The theoretical implications of the present research are discussed in detail in Chapter 4 (the General Discussion).

Methodological Issues: Indexing Subitizing, Counting and ANS

My initial question was whether different kinds of nonsymbolic quantitative tasks (identification, comparison, and ordering) from the same range (small, medium, and large) tapped the same underlying constructs (i.e., the subitizing, counting, and approximate number systems, respectively). Attempts to create latent factors reflecting this organization were unsuccessful, probably due to the fact that correlations among tasks expected to tap the same constructs were weak or insignificant. There were several potential methodological reasons for why this might be. To try to disentangle task-type from range (and perhaps quantitative system) the present project included some tasks that were novel—or at least novel to some ranges. The fact that they were novel means that their psychometric properties (i.e., convergent and discriminant validity) were not known prior to conducting this research.

In a recent paper, Dietrich et al. (2015b; see also Inglis & Gilmore, 2014; Price et al., 2012) clearly illustrate how different aspects of task methodologies can impact task reliability and validity. Their paper focused on ANS tasks, as the ANS literature is substantial enough to permit this kind of methodological overview. While this same analysis is not possible with subitizing and counting tasks, it is my view that the methodological issues raised in Dietrich et al.'s paper come to bear on the present research as well. ANS tasks have demonstrated poor convergent validity—that is ANS tasks are not always correlated with each other. Price et al. (2012) examined three dot comparison tasks in three formats: paired, sequential, and intermixed presentations. They found that paired-comparison performance (numerical ratio effect [NRE]) was only moderately correlated (r = .39) with intermixed and sequential comparison formats (r = .50); intermixed and sequential comparison tasks were correlated r = .68. Thus, performance on these dot comparison tasks was not as highly correlated as might be expected, suggesting that variations within the same *type* of task (nonsymbolic comparison, in this case) might have an impact on results.

It is possible that task variability also affected small nonsymbolic identification in the present research. Both the identification and rapid identification of small quantities involved naming the quantities 1 to 3, however, they were only correlated *r* = .27. Note that the identification task involved identifying quantities 1 to 9, and the RTs for 1 to 3 were isolated afterwards to create the nonsymbolic identification of small quantities performance indicator. In contrast, the *rapid* identification task involved 1 to 3 dots only. From a face validity perspective, a measure of small identification derived from a task that includes small quantities *only* (as in the rapid identification measure) seems like a purer measure of subitizing than one derived from the task with 1 to 9 dots. Perhaps enumerating 1 to 9 dots involves switch costs as participants switch from accessing subitizing and counting systems.

Dietrich et al. (2015b) argue that task objective also has an impact. Paired-comparison ANS tasks require participants to select the greater quantity, whereas same-different tasks require them to indicate whether the two quantities are the same or different. Smets et al. (2015) found that same-different performance was not significantly correlated with paired-comparison performance (*r* = .15 for accuracy and *r* = -.28 for Weber fraction). In contrast, Sasanguie et al. (2011) observed a significant correlation between paired-comparison and same-different performance (numerical distance effect [NDE]; *r* = .52.). Different task demands in the present research perhaps explain the weak correlations among large quantity identification and large quantity comparison and ordering.

This task-goal issue could also explain why nonsymbolic comparison of small quantities was only weakly correlated with small identification tasks (*r* = .26 for both)—even though both were expected to activate the subitizing system. Note that there was also evidence that measures were more strongly related when task demands were the same (e.g., nonsymbolic comparison in the small and medium range correlated $r = .65$; ordering in the small-medium and large ranges correlated $r = .64$); this pattern was true for symbolic tasks as well. To summarize, tasks within the same quantitative range, and presumed to index the same underlying constructs, were not highly correlated, demonstrating poor convergent validity. This lack of validity could be because the tasks tap different underlying constructs or because their unshared variance is attributed to different domain-general cognitive skills (Gilmore et al., 2014; Price et al., 2012).

Another threat to convergent validity within the ANS literature stems from the fact that the exact same ANS task can be (and has been) scored in multiple ways, and, unfortunately these different performance indicators (labeled dependent variables in Dietrich et al., 2015b) have not demonstrated high correlations with each other (Price et al., 2012; Inglis & Gilmore, 2014). These performance indicators include accuracy (with or without response times [RTs]), the Weber fraction, the numerical distance effect (NDE), and the numerical ratio effect (NRE; refer to Dietrich et al., 2015b for a detailed description of these measures). The Weber fraction reflects the minimum ratio which can be discriminated, beyond which participants perform at chance; lower values indicate better performance. As described previously, the ANS is an approximate, ratio-dependent system. That is, the closer the quantities are to one another (in ratio and distance), the more difficult it is to discriminate between them, and this is evidenced in accuracy and RTs. Thus, the NRE and NDE reflect individual differences in the "costs" associated with more difficult comparison trials (those trials with closer quantities and/or ratios relatively closer to 1:1). It is assumed that smaller NREs and NDEs reflect more precise ANS representations. The NRE is recommended over the NDE because the NRE is believed to better reflect the magnitude of the numerosities, and hence ANS precision (Dietrich et al., 2015b).

Although accuracy and Weber fractions were highly correlated in the literature (presumably because the Weber fraction includes accuracy in its calculation), NRE based on accuracy and RT data were not highly correlated with accuracy (NRE_{acc} r = .14 and NRE_{RT} r = .44) or Weber fraction (NRE_{acc} r = .40 and NRE_{RT} r = .37; Price et al., 2012; Inglis & Gilmore, 2014), raising concerns about the convergent validity of different performance measures derived from the *same* ANS tasks (Dietrich et al., 2015b). Reflected in these findings is the fact that RT and accuracy information relate to other variables in different ways (e.g., refer to NRE accuracy versus RT correlations above). Dietrich et al. (2015b) conclude that "it is unclear whether ANS theory would predict a relationship between accuracy and RT based measures" (p. 4, Dietrich et al., 2015b; see also Inglis & Gilmore, 2014).

The ANS findings presented above are described to illustrate the following: Even if task demands are the *same*, the different indicators of performance are not necessarily related. I presume that this would only be intensified when different kinds of tasks intended to tap the same construct are scored differently—as was the case in the current research. Because of the way certain tasks were administered, performance measures were accuracies, *P*, or mean percent absolute error. The different performance scoring methods might explain the weak or lacking correlations observed.

In the present study (and consistent with past research) the comparison and ordering tasks required participants to select one of two responses (and thus were correct or not); accuracies and RTs were recorded. If accuracy was greater than 85%, a corrected

RT (*P*) was calculated. However, accuracy was less than 85% for nonsymbolic comparison large (ANS), and nonsymbolic ordering small-medium and large; performance was simply accuracy for these measures. Rather than a two-choice response, nonsymbolic identification required participants to name (or guess) the number of dots shown; this was an open response with a voice key RT. Accuracy was 100% for small identification (subitizing), so RT was the performance indicator. Exact accuracy in the medium (counting) and large (ANS) identification tasks was not expected, and instead mean percent absolute error scores were calculated. RTs were corrected by the mean percent absolute error to score the counting task, as percent absolute error was not very large, but the counting task was missing a substantial number of RTs. The RT performance measure was retained as it was consistent with past counting tasks, however the loss of RTs may have decreased the construct validity of the measure.

Another issue with the medium identification (counting) task was that stimuli presentation time was limited to make it consistent with the presentation time for comparison and ordering tasks. Upon reflection, the stimuli presentation time (and hence, voice-key activation period) should have been unlimited or at least extended. An unlimited presentation time would be consistent with past enumeration tasks, and would ensure participants had enough time to count, and were not guessing—which might tap the ANS. Stimuli presentation time was limited in the large nonsymbolic identification task as well. A larger proportion of RTs were missing as participants did not respond in the time allotted, but in this case the short presentation time was unlikely to affect the construct being measured. Performance on the nonsymbolic identification of large quantities was mean percent absolute error. To summarize, the performance measures for nonsymbolic quantitative tasks by range were as follows: a) Subitizing (small): All error-corrected speed/RT-based. All three were significantly, if weakly, correlated; b) Counting (medium): *P*s for identification and comparison, and accuracy for ordering. Ordering performance did not correlate with the other two measures; and c) ANS (large): Mean percent absolute error for identification, accuracy for both comparison and ordering. Identification performance did not correlate with the other two measures. This pattern suggests that task scoring (as well as task demands) had an impact on item correlations.

Another issue identified within the ANS literature is a lack of reliability within some ANS tasks. NRE and NDE performance scores are associated with poor test-retest (Inglis & Gilmore, 2014) and split-half reliabilities (Sasanguie et al., 2011). Poor reliability within tasks can contribute to a lack of convergent validity, as reliability sets an upper bound on correlations with other variables (Goodwin & Leech, 2006). Parallel-form (rapid identification only) and split-half reliabilities for all quantitative tasks were good to adequate, ranging from .63 to .97 (only symbolic comparison medium and nonsymbolic comparison medium were less than .70). In sum, poor reliabilities do not seem to explain the weak correlations among measures intended to tap the same underlying constructs in the present research.

To summarize, it was hypothesized the nonsymbolic quantitative measures from different tasks (identification, comparison, and ordering) but from within the same range (small [1 to 3], medium [5 to 9], and large [10 to 100]) would load together to create subitizing, counting, and ANS factors, respectively. This hypothesis was not supported, however. Three possible explanations for the lack of support for the theorized factor structure are as follows: 1) Identification, comparison, and ordering measures within the same range simply do not tap the same underlying constructs, 2) Measures tap the same underlying constructs, but also include substantial variance which is systematically related to other cognitive processes by task-type (an issue of discriminant validity), 3) Measures tap the same underlying constructs, but different scoring methods contribute to poor convergent validity among tasks. It is important to note that two tasks might tap the same underlying construct, while the third does not. For example, identification of large quantities might not access the ANS. Izard and Dehaene (2008) claim that dot estimation with large quantities should access the ANS, but there is no support for this claim because, as far as I can tell, no one has looked at large identification performance in relation to traditional ANS comparison tasks or mathematics performance.

This discussion of methodological concerns highlights the importance of considering the psychometric properties and scoring procedures of research tasks. If a new task is developed, it should ideally be examined along with traditional tests to assess its validity. If the new and traditional measures tap the same construct, they should correlate with each other and other related constructs (convergent validity). The discriminant validity of the tasks can also be compared (e.g., How well do they predict the outcome?; How "pure" are they [are they tapping other constructs]?). The internal reliability of the new task should also be assessed, along with test-retest reliability (at some point, if not right away; Dietrich et al., 2015b). Any observed effects (or lack of effects) associated with the new task can also be compared to those observed with the traditional task—which can be helpful if spurious results are obtained. If effects are obtained, replicability is required before any broad conclusions drawn. The implications of scoring the same task in different ways should also be explored.

In the present study I chose to include error-corrected RTs (*P*) as often as possible, as this seemed like a reasonable way to capture the two performance aspects (accuracy and RT) in one measure. *P* was not calculated when accuracy was less than 85%, as the RT correction would become extreme with high error rates. In those cases there was enough variability in accuracy to meaningfully reflect individual differences in performance (i.e., with no variability there is no predictive power; scores at ceiling are undesirable). Future research might explore the different data patterns associated with scoring tasks in different ways (as was done in Dietrich et al., 2015b), as well as potential tasks to index the subitizing and counting systems. With reliable and valid measures in place, researchers could continue to explore subitizing and counting in relation to mathematical and number line skill among adults. In the present study, single, observed variables were selected as measures of the subitizing, counting, and ANS systems. The counting and ANS measures were consistent with those employed in past research. The subitizing measure was consistent with past research in the sense that it required identification of small quantities, however it differed from past measures in that it was not derived from an enumeration task with quantities in both the subitizing and counting ranges. This task demonstrated predictive validity and parallel-form reliability in Studies 1 and 2, and thus seemed a reasonable choice.

Relations among quantitative skills: Quantitative-Only versus Full Model SEMs

Prior to conducting the Full SEMs (which included domain-general cognitive skills and experiential variables), the quantitative variables were examined in relation to each other and in relation to the mathematical or number line outcome. Hypothesized paths between quantitative skills were modeled. As hypothesized, subitizing predicted counting (H3), and symbolic identification and comparison (H4). This finding is consistent with the view that the subitizing system scaffolds the acquisition of the counting system and subsequent symbolic number system knowledge. Counting was also predictive of both symbolic identification and comparison, supporting Hypothesis 7. However, executive function turned out to be a strong predictor of subitizing, counting, symbolic identification, and symbolic comparison, and once it was added in the Full SEM, subitizing no longer predicted counting, symbolic identification, or symbolic comparison. Relations between counting and symbolic identification and comparison were also reduced. It seems that the variance executive function shared with subitizing overlapped completely with the variance shared by subitizing and the other quantitative variables (counting, symbolic identification, and comparison). This is perhaps not surprising considering that executive function was more strongly correlated with counting, symbolic identification, and symbolic comparison than was subitizing, and subitizing and executive function were highly correlated as well.

The other relations among quantitative variables were consistent in both the Quantitative-Only and Full SEMs. ANS was a significant predictor of nonsymbolic ordering large, presumably because both tasks tap ANS skill. Symbolic comparison and ANS (nonsymbolic comparison large) were modeled to be correlated, as were symbolic and nonsymbolic ordering, as these tasks shared the same task demands; these correlations were significant. Symbolic identification was not a significant predictor of symbolic comparison in either the Quantitative-Only or Full SEMs. Symbolic comparison was a strong predictor of symbolic ordering across SEMs, as well.

Predictors of Mathematical Outcomes: Quantitative-Only versus Full Model SEMs

To test the relative contribution of the different kinds of quantitative skills, subitizing, counting, ANS, nonsymbolic ordering large, and symbolic ordering were entered as predictors of each mathematical or number line outcome. The symbolic ordering task was selected to represent the symbolic quantitative system. The results for the Quantitative-Only SEMs were quite similar for all of the mathematical outcomes (those requiring exact calculation): Symbolic ordering was a significant predictor, while nonsymbolic ordering, ANS, counting, and subitizing were not. Thus, symbolic ordering appeared to mediate the relations between other quantitative measures and mathematical performance. There were two exceptions to this pattern: 1) Subitizing was also a significant unique predictor of arithmetic fluency (consistent with Studies 1 and 2), and 2) Symbolic comparison was also a unique predictor of speeded addition.

The number line Quantitative-Only SEMs showed a different pattern; ANS did not predict the other mathematical outcomes but did predict number line performance. Like the mathematical outcomes, symbolic number line was predicted by symbolic ordering; subitizing, counting, and symbolic comparison were indirectly related. Nonsymbolic ordering of large quantities had a marginally significant direct effect, and ANS a significant total effect, on symbolic number line. Unlike all of the other outcomes, the nonsymbolic number line was not directly predicted by symbolic ordering (or any other symbolic skill), or indirectly by any other quantitative measure. (Note that since symbolic ordering was not predictive, there was no significant pathway through which subitizing, counting, or symbolic comparison could have an indirect effect). The nonsymbolic number line was instead predicted by ANS, nonsymbolic ordering, and counting (marginal).

As described above, adding executive function to the Full SEMs changed the interrelations *among* the quantitative measures. Because subitizing no longer accounted for unique variance in counting, symbolic identification, and symbolic comparison, it could no longer have a significant indirect effect on any outcome, as it did in the Quantitative-Only models. Similarly, executive function, counting, symbolic identification, and symbolic comparison shared overlapping variance, which resulted in a decrease in the strength of counting as a predictor of symbolic identification and comparison. This left little room for counting to have an indirect effect on outcomes. To summarize, one of the main differences between the Quantitative-Only and Full SEMs was that subitizing and counting no longer exhibited indirect effects on the outcomes; these indirect effects were essentially replaced by the indirect effects of executive function—which were significant for all outcomes except calculation knowledge. This means that it was difficult to separate the executive function factor from the subitizing and counting measures (subitizing in particular). Basically, in the Full SEMs executive function accounted for all of the variance in quantitative skills that was attributed to subitizing in the Quantitative-Only model. Executive function accounted for the same variance subitizing had, plus more, as the total variance accounted for in quantitative measures increased in Full SEMs relative to Quantitative-Only SEMs (refer to Table 3.16).

In general, the same quantitative measures that directly predicted the mathematical outcomes in the Quantitative-Only SEMs remained significant direct predictors in the Full SEMs, however regression coefficients tended to be smaller. The one exception was that symbolic ordering was no longer a significant predictor of speeded addition (symbolic comparison was). Subitizing remained a significant predictor of arithmetic fluency. Symbolic ordering remained a significant predictor of arithmetic fluency, calculation knowledge, and symbolic number line. Nonsymbolic ordering large was a significant predictor of nonsymbolic number line, and a marginally significant predictor of symbolic number line. ANS and counting (marginal) predicted the nonsymbolic number line only. Consistent with the Quantitative-Only models, symbolic comparison exhibited small but significant indirect effects on arithmetic fluency and calculation knowledge, although these effects were also decreased in the Full SEMs.

In the Full SEMs, the decreases in the direct effect of symbolic ordering and indirect effect of symbolic comparison on arithmetic fluency, speeded addition, and calculation knowledge were due to the inclusion of the experiential variables, and not due to the inclusion of executive function. Since executive function was not specified as a direct predictor of these calculation-based outcomes, it could not detract from the variance accounted for by symbolic ordering or comparison. Executive function was specified to predict symbolic comparison and indirectly predict symbolic ordering; thus, in terms of accounting for variance, it worked with—rather than against—the predictive strength of these factors. Note that perceived math competence (PMC) predicted arithmetic fluency, speeded addition, and calculation knowledge, and was also correlated with executive function, symbolic comparison, and symbolic ordering. This pattern suggests that PMC was responsible for the decreased effects of symbolic comparison and ordering in the Full SEMs relative to the Quantitative-Only ones. Thus, participants' experiences accounted for unique variance in the outcomes, whereas domain-general skills did not. Executive function had an indirect effect by predicting the quantitative skills, which in turn predicted mathematical outcomes. The number line results also support this view: PMC did not predict symbolic number line performance, and the direct and indirect effects did not change for this outcome (i.e., there was no variance in symbolic number line that overlapped with both symbolic ordering and PMC). The nonsymbolic number line was predicted by PMC, and by counting, ANS, and nonsymbolic ordering. The coefficients for these quantitative predictors remained the same size in the full SEM, but they were not correlated with PMC. In sum, PMC was the variable that decreased the predictive strength of the symbolic ordering coefficient.

Whether the changes in the direct and indirect effects of symbolic comparison and ordering were due to executive function or experiential variables was formally tested. This was done by creating intermediate SEM models in which executive function was added to the Quantitative-Only model prior to adding in the experiential variables. In these intermediate models, the regression coefficients, indirect effects of symbolic comparison, and variance accounted for in the mathematical outcomes were the same as in the Quantitative-Only models. It was not until the experiential variables were added to the SEMs that the regression coefficients decreased, and the variance accounted for in the mathematical outcomes increased. This demonstrates that PMC, age, and education location were responsible for unique variance in the mathematical outcomes in the Full SEMs. PMC was likely responsible for the change in symbolic ordering paths as it was the only experiential variable that shared variance with both symbolic ordering and the outcomes.

Summary

Table 3.17 summarizes the supported and unsupported hypotheses for Study 3. In this study, the subitizing, counting, ANS, and symbolic quantitative systems were measured and examined in relation to each other and in relation to different kinds of mathematical and number line outcomes. Domain-general cognitive abilities (executive function) and experiential variables were also taken into consideration in the final structural equation models. Nonsymbolic quantitative tasks were expected to group together by range (small, medium and large) to form factors representing the subitizing, counting, and ANS respectively. However, the different kinds of tasks (identification, comparison, and ordering tasks) within the same range were not well-correlated and the hypothesized factor structure did not fit the data. This could be because different tasks within a range do not tap the same underlying quantitative system (e.g., small range [1 to 3 dots] and the subitizing system], or it could be due to unexplored methodological issues (e.g., task administration or scoring).

Executive function predicted subitizing, counting, symbolic identification, and symbolic comparison. Much of the overlapping variance among subitizing and counting, symbolic identification, and symbolic comparison overlapped with the variance these variables shared with executive function, and subitizing no longer predicted them once executive function was taken into consideration. Subitizing was a direct predictor of arithmetic fluency only. Counting had little direct or indirect effect once executive function was taken into consideration, but it was a marginal direct predictor of nonsymbolic number line performance. Not surprisingly, the symbolic quantitative system was the best predictor of all of the outcomes except for nonsymbolic number line performance. The ANS was not predictive of any of the outcomes requiring exact calculation, but it was related to number line performance (moreso for nonsymbolic than symbolic). Experiential variables were also found to account for unique variance in the mathematical and number line outcomes. The theoretical implications of the present findings, along with the findings of Studies 1 and 2, are discussed in depth in the General Discussion in Chapter 4.

Table 3.17. *Summary of Supported Hypotheses in Study 3*

Category	Hypotheses	
NS small (1 – 3): **Subitizing**	1. Subitizing skill will not relate to ANS skill	✓
	3. Subitizing skill will predict counting skill ➢ Subitizing *not* predictive of counting in full SEM once executive function was included as a predictor of quantitative skills	QO
	4. Subitizing skill will predict symbolic identification and comparison ➢ Subitizing was *not* predictive of symbolic identification and comparison in full SEM once executive function was included as a predictor of quantitative skills	QO
NS large (10 - 100): **ANS**	5. ANS skill will not uniquely predict *exact* mathematical calculation ➢ ANS did not predict arithmetic fluency, speeded addition, or calculation knowledge	✓
	6. ANS skill will predict approximate numerical outcomes, (i.e., number line performance) ➢ ANS (total effect) and nonsymbolic ordering large predicted symbolic number line ➢ ANS and nonsymbolic ordering large predicted nonsymbolic number line	✓
NS med (5 - 9): **Counting**	7. Counting will predict symbolic identification and comparison ➢ Effect of counting decreased by presence of executive function in full SEMs, but small effect still present	✓
Task-type and format	8. Nonsymbolic identification, comparison, and ordering tasks from the same range (small, medium, and large) will tap the same underlying constructs (subitizing, counting, & ANS) ➢ Variables from the same range but different tasks were not well-correlated ➢ The data did not fit the theorized small, medium, and large factor structure	X
	9. Less complex symbolic quantitative skills will predict more complex ones ➢ Symbolic identification *did not* predict symbolic comparison ➢ Symbolic comparison *did* predict symbolic ordering	P

Note: In terms of the hypotheses, ✓ = supported, X = not supported, P = partially supported. QO = Quantitative-Only SEM. NS = Nonsymbolic. ANS = Approximate number system.

Table 3.17. *Summary of Supported Hypotheses in Study 3 -continued-*

Category	Hypotheses	
Task-type and format	10. Nonsymbolic quantitative skills will not predict calculation-based mathematical outcomes beyond symbolic quantitative skills ➢ Arithmetic fluency: *No*; symbolic ordering uniquely predicted, but so did subitizing ➢ Speeded addition: *Yes*; symbolic comparison uniquely predicted ➢ Calculation knowledge: *Yes*; symbolic ordering uniquely predicted	P
	11. Symbolic ordering will be the best quantitative predictor of calculation-based mathematical outcomes ➢ *Yes* for Arithmetic fluency, calculation knowledge, and symbolic number line ➢ *No* for Speeded Addition; symbolic comparison uniquely predicted instead ➢ No for nonsymbolic number line (but was not expected for this outcome) ➢ Symbolic comparison also had a significant indirect effect on arithmetic fluency, calculation knowledge, and symbolic number line	P
Domain-general cognitive skills	12. Domain-general cognitive skills will predict domain-specific quantitative skills and have indirect effects on mathematical outcomes ➢ Executive function factor significantly predicted subitizing, counting, symbolic identification, and symbolic ordering ➢ Executive function had significant indirect effects on all outcomes except calculation knowledge	✓
Experience	13. Experiential factors will uniquely predict mathematical outcomes ➢ Perceived math competence (PMC) predicted all outcomes except symbolic number line ➢ Age predicted all calculation-based mathematical outcomes, but not number line tasks ➢ Education location: Asian-educated participants performed significantly better on arithmetic fluency, speeded addition, and calculation knowledge measures than those educated elsewhere	✓

Note: In terms of the hypotheses, ✓ = supported, X = not supported, P = partially supported. QO = Quantitative-Only SEM. NS = Nonsymbolic. ANS = Approximate number system.

CHAPTER 4: GENERAL DISCUSSION

The current dissertation explored domain-specific quantitative skills as unique predictors of adults' mathematical outcomes, while simultaneously considering experiential factors and executive function (i.e., domain-general cognitive abilities). Theoretically-relevant quantitative constructs were defined and assessed, namely: subitizing, counting, approximate number system (ANS), and symbolic quantitative skills. Both the subitizing and approximate number systems are presumed to be core quantitative systems, given the evidence that human infants and other animals, such as monkeys can a) keep exact track of small sets of objects (i.e., subitize) and b) can distinguish between large quantities, assuming the quantities are large enough and within the discriminable ratio-limit (an ability attributed to the approximate number system; see Feigenson et al., 2004 for a review). People in literate cultures learn a symbolic quantitative system; this system permits *exact* representation and manipulation of quantities beyond the subitizing range—and often beyond the range that is easily comprehended. But how do humans learn these symbolic quantity representations? This question is known as the "symbol-grounding problem" (Leibovich & Ansari, 2016). Learning how to count is the first step in acquiring the symbolic quantitative system.

The present research was framed by three theoretical accounts. Each account posits how symbolic quantity representations relate to nonsymbolic quantity representations, and each account emphasizes the importance of one of the subitizing, counting, or approximate number systems in relation to later mathematic success. In this General Discussion, I first briefly summarize these three accounts and the evidence to support them. Second, I discuss how the current research speaks to these accounts. Third, I discuss the contribution of symbolic versus nonsymbolic quantitative skills when predicting mathematical outcomes. Fourth, I discuss the role of domain-general cognitive abilities, and fifth, the role of experiential factors. The research limitations, implications, and future recommendations are also discussed.

The Three Accounts

ANS-is-key account. According to this account, the ANS provides a sophisticated representation of quantity; the challenge is simply learning the symbols that map to the nonsymbolic representation already in place (Dehaene, 1997; Gallistel & Gelman, 1992; Piazza et al., 2010). According to this view, the ANS might scaffold the development of counting abilities, and should relate to symbolic skill and mathematical performance (once developed). As described in Chapter 1, two recent papers review the data and conclude that the current evidence does not support this account (Leibovich & Ansari, 2016; Lyons & Ansari, 2015). ANS skills do not appear to relate to the development of counting skills (Odic et al., 2015) nor to mediate relations between symbolic quantitative skills and mathematical outcomes (Fazio et al., 2014; Sasanguie et al., 2014). In fact, symbolic number skills may actually predict ANS performance (Mussolin et al., 2014). The correlation between ANS acuity and symbolic arithmetic performance among adults and children (Chen & Li, 2014; De Smedt et al., 2013, Fazio et al., 2014; Schneider et al., 2016) is the only empirical support for this account—however, correlation is not definitive proof that the ANS is key to mathematical success.

Subitizing-is-key and Subitzing+Counting-is-key. According to both of these accounts the subitizing system—which permits quick and exact enumeration (or identification) of up to 3 or 4 quantities without counting—is crucial to the development of the symbolic system and later mathematical skill (Carey, 2001, 2004; Landerl et al., 2004; Starkey & Cooper, 1980). On this view, children have representations for these small quantities; they must simply learn the counting words (symbols) and how they map to these known quantities (Carey, 2001, 2004). According to Carey, once some of the basic counting principles are understood within this set, children generalize them to greater quantities. The subitizing-is-key and subitizing+counting-is-key accounts differ in the emphasis placed on the subitizing versus counting systems. According to the subitizing-is-key account, individual differences in subitizing should predict not only counting, but also symbolic number systems skills and arithmetic performance. In contrast, in the subitizing+counting-is-key account put forth by Noël and Rousselle (2011), once the counting system has been established, individual differences in counting should be what predicts later math success (not subitizing, per se). Individual differences in subitizing have been found to predict later numerical skill among children (e.g., Hannula-Sormunen et al., 2015), however, not much research has explored the role of subitizing across development. By comparison, many studies have shown that individual differences in counting are related to math achievement among children—but few studies have included both subitizing and counting, so their relative contributions to mathematical skill are not known. To my knowledge, no prior study has examined individual differences in subitizing or counting in relation to mathematical achievement among adults.

The Current Research: Evaluating the Three Accounts

The relative contribution of the ANS to mathematical skill. Consistent with the view that the ANS aids the development of the symbolic quantitative system, ANS and symbolic quantitative skills were correlated in Studies 1 and 3 of the current thesis—however, as described in Leibovich and Ansari (2016) the causal direction of the correlation is unknown, and thus no path was specified between ANS (nonsymbolic comparison) and symbolic comparison in Study 3. Instead these constructs were set to be correlated (as represented by the double-headed arrow). ANS performance was expected to correlate with mathematical outcomes (e.g., Schneider et al., 2016). In Study 1, individual differences in the ANS were weakly related to arithmetic fluency, but did not uniquely predict when subitizing was also entered as a predictor. In Study 3, the traditional ANS measure (nonsymbolic comparison of large quantities) was not significantly correlated with any of the mathematical outcomes requiring exact computation, that is, arithmetic fluency, speeded addition, or calculation knowledge (BMA). This finding is not consistent with recent meta-analyses that estimate a correlation between ANS acuity and arithmetic performance in the .20 to .24 range (Chen & Li, 2014; Fazio et al., 2014; Schneider et al., 2016), however it is consistent with the fact that some past research has failed to find relations between ANS and mathematical outcomes. An explanation for the lack of correlation between ANS and math performance in Study 3 was not found.

Although Schneider et al. (2016) included several possible moderators of the relation between ANS and math performance—including ANS task scoring, the measure of mathematical competence, and poor reliability--none of these were found to account for the lack of relations between the ANS and mathematical performance in the present research. The ANS tasks in Studies 1 and 3 differed slightly. The ANS measure in Study 1 was the Panamath, which relative to Study 3 included smaller numerosities. The ANS task in Study 3 included larger quantities relative to past research (from 10 to 100), and difficult ratios (up to 11:10), however, there was no evidence that the ANS task was too difficult. Overall accuracy was 77% and no single ratio was below 65% accuracy—thus participants seemed to be performing above chance. The current ANS task was designed to be consistent with many of the recommendations put forth by Dietrich et al. (2015b). In all three studies, the visual properties of the dot stimuli were controlled across the task (e.g., area was positively correlated for half of trials and perimeter controlled for the half; dots varied in size); stimuli were simultaneously presented in pairs. The goal was to identify the greater quantity, dots were well above the subitizing range, and performance was scored as accuracy (not numerical ratio effect [NRE] or Weber fraction). In Study 3, there were a large number of trials (180) and split-half reliability of the ANS task was good (.80), so the lack of correlation is not due to poor reliability limiting the size of the potential correlation (Goodwin & Leech, 2006). The reliability of the arithmetic fluency measure was also good. Perhaps the arithmetic fluency measure used in the current research is simply less related to ANS performance than other mathematical performance measures. It is interesting to note that the correlations between ANS and arithmetic fluency in Studies 1 and 3 were similar (.16 versus .12) suggesting consistency within the present research.

I am reluctant to call the nonsymbolic comparison (1 to 9) measure in Study 2 a measure of the ANS, because I believe it also taps subitizing and/or counting skill, it was consistent with “ANS” measures employed in other research (e.g, Lyons & Beilock, 2011). In Study 2, nonsymbolic comparison (1 to 9 dots) and arithmetic fluency were correlated at .43; this correlation is well above the expected correlation ($r = .24$) reported by Schneider et al. (2016)—presumably because the present ANS measure included smaller numerosities. These findings suggest that the distribution of quantities included in ANS tasks may be an important moderator of the relation between the task and math outcomes. This finding is in-line with Dietrich et al.’s (2015b) recommendations and the current theoretical perspective. Researchers should ensure that ANS measures do not include small quantities as they may tap other quantitative systems (subitizing and/or counting), thus creating a lack of distinctiveness across tasks and leading to results that are difficult to interpret in terms of the relative importance of the different quantitative systems.

Study 3 also included nonsymbolic ordering of large quantities, which may also tap the ANS; it shares stimuli and presumably requires sequential comparisons based on approximate quantities. This task is perhaps more difficult or sophisticated than the traditional ANS task because it also requires comparison, but comparisons are conducted among three, rather than two, quantities. In Study 3, performance on nonsymbolic ordering of large quantities was significantly correlated with mathematical outcomes requiring mathematical computation; these correlations were close in magnitude to the ANS-math performance correlations expected based on meta-analytic results (Chen & Li, 2014, $r = .20$; Schneider et al., 2016, $r = .24$). Nonsymbolic ordering was correlated .19 with arithmetic fluency, .17 with speeded addition, and .20 with calculation knowledge ($N = 149$ to 153; ps $< .05$). The ANS comparison and nonsymbolic ordering tasks were moderately correlated, which is some evidence of convergent validity (i.e., that these tasks tap the same underlying construct).

It is my opinion that there is no logical reason for the ANS—an approximate quantity representation—to predict outcomes requiring *exact* symbolic mathematical calculation. Why employ an approximate system for exact calculation—especially when an exact, symbolic system (a tool in itself) has been developed and learned for this purpose? However it does seem reasonable for ANS acuity to predict outcomes requiring approximate or estimated responses, such as number line tasks. Consistent with this view, ANS (comparison) and nonsymbolic ordering were found to relate to both symbolic and nonsymbolic number line performance, with correlations above .30.

The relative contribution of the subitizing system to mathematical skill. The findings of the present research suggest that subitizing is a relevant quantitative system that has been largely ignored. To my knowledge, the present research was the first to assess subitizing as a predictor of arithmetic fluency in adults. In Study 3, subitizing was moderately correlated with counting (whereas ANS was not correlated with counting), as well as with symbolic identification, symbolic comparison (Studies 2 and 3) and speeded addition. These correlations are consistent with the view that the subitizing system scaffolds the acquisition of symbolic quantitative representations and with the view that the *exact* nature of the subitizing system makes it a better candidate than the ANS to scaffold the development of a symbolic number system (Butterworth, 2010; Carey, 2001). Subitizing was also found to uniquely predict arithmetic fluency across all three studies. This relation persisted even when ANS (Studies 1 and 3), counting (Study 3), symbolic quantitative skills (Studies 2 and 3) and experiential factors (all three studies) were taken into account, suggesting that the pattern of correlation between subitizing and arithmetic fluency is stable among adults. Similar correlations have been reported for children (Hannula-Sormunen et al., 2015; LeFevre et al., 2010; vanMarle et al., 2014).

It was surprising to find that individual differences in such a basic quantitative process accounted for unique variance in arithmetic fluency among adults. I had anticipated that symbolic skills (such as number comparison or order judgments) would mediate relations between subitizing and arithmetic performance. Subitizing was not a unique predictor of any other mathematical outcome, however. These findings suggest that much more attention be paid to individual differences in the subitizing system and how they might predict specific mathematical performance across development. Notably, subitizing was highly related to executive function in Study 3; this is discussed below, in the section on domain-general cognitive abilities. Arithmetic processes also involve executive functions, particularly working memory (DeStefano & LeFevre, 2004) and inhibitory skills (LeFevre, Bisanz, & Mrkonjic, 1988). These relations may account for the link between subitizing and arithmetic fluency.

The relative contribution of the counting system. Study 3 was the only study to include a counting measure. In the most conservative interpretation of the subitizing+counting-is-key account, counting skill would fully mediate relations between subitizing and mathematical outcomes. In actuality, subitizing was a significant direct predictor of arithmetic fluency in Study 3. In contrast, counting did not directly predict any mathematical outcome, except for a weak relation (i.e., marginally significant) with the nonsymbolic number line. Counting was a significant predictor of symbolic identification and comparison, but these relations did not translate into significant indirect effects on outcomes when executive function was also included in the model (relations with domain-general skills are discussed below). Thus, although counting was related to other quantitative skills and to mathematical outcomes, it was not uniquely predictive beyond symbolic, domain-general, and experiential variables. This finding does not necessarily mean that counting is not important to the development of the symbolic representations and mathematics, especially given the substantial evidence that counting skills are predictive of math achievement among children. Instead, it suggests that other quantitative skills are more important for predicting math in adult populations. This finding also requires replication, as little to no research has examined counting in relation to other quantitative skills or mathematical performance among adults. Furthermore the counting measure used in Study 3 was missing a substantial number of response times due to a restricted time limit for responses. Although the measure was still reliable, its construct validity might be reduced relative to counting measures without time limits.

Summary. When contrasting the relative importance of individual differences in the subitizing, counting, and approximate number systems to mathematical performance, I found that subitizing consistently accounted for unique variance in arithmetic fluency performance (all three studies) and the ANS was uniquely predictive of number line performance (Study 3). It seems plausible that the subitizing system scaffolds the acquisition of symbolic number skills.

Symbolic versus Nonsymbolic Quantitative Skills

In Study 3, symbolic quantitative skills were organized by task, creating identification (naming), comparison, and ordering (order judgment) factors. In Studies 2 and 3, symbolic identification was not predictive of symbolic comparison (and not related to much else). These findings are consistent with the evidence that numerals are processed asemantically in adults (Dehaene & Akhavein, 1995; Myers & Szűcs, 2015; Reeve et al., 2012)—that is, simply naming numerals does not seem to require access to their meaning. Conversely, symbolic comparison and order judgments presumably require access to quantity and order representations. I interpret this as evidence that identification did not require access to the same underlying cognitive processes as comparison and ordering.

In Study 3, the symbolic comparison and ordering factors were highly correlated. For most of the SEMs in Study 3, relations between symbolic comparison and math outcomes were mediated by symbolic ordering with the exception of speeded addition, which was directly predicted by symbolic comparison. The recent meta-analysis by Schneider et al. (2016) clearly shows that symbolic quantitative skills are more strongly related to math performance than nonsymbolic ones (see also De Smedt et al., 2013). The results of the present research are mostly consistent these findings in that symbolic quantitative processes stood out as strong predictors of mathematical outcomes. Unexpectedly, subitizing uniquely predicted arithmetic fluency beyond symbolic quantitative skills in Studies 2 and 3. ANS performance (but not symbolic quantitative skills) predicted nonsymbolic number line performance, presumably because both tasks involve comparisons of quantities. Thus, in general, across all three studies, symbolic number skills (number comparison or order judgments) were the strongest quantitative predictors of mathematical outcomes presented in symbolic format.

The symbolic order judgment task is a relatively novel task in comparison to the symbolic comparison tasks (Lyons & Beilock, 2009, 2011; Lyons et al., 2013). The order judgment task may tap a similar, but more sophisticated or nuanced quantity representation. Alternatively, it may capture additional representational features of symbolic digits, specifically, their relative positions in the counting sequence. Symbolic number comparisons, at least hypothetically, only require quantity information, not relative order. It is worth noting that symbolic comparison would be a strong predictor of mathematical outcomes if order judgment was not included in the analyses; the correlations these variables shared with all other model variables were very similar. Further research is needed to disentangle the specific features of symbolic comparison and symbolic order tasks that are predictive of mathematical outcomes.

Domain-general cognitive skills

In Study 3, five domain-general cognitive abilities were combined to create a latent executive function factor. Domain-general cognitive skills were included in Study 3 because past research has shown that such skills predict arithmetic performance for children and adults (e.g., DeStefano & LeFevre, 2004; Raghubar et al., 2010; Szűcs et al., 2014). Executive functioning was hypothesized to directly predict quantitative skills, and indirectly predict mathematical outcomes through the quantitative skills (i.e., quantitative skills would mediate relations between executive function and mathematical outcomes). These hypotheses were supported. Executive function was a strong predictor of subitizing, counting, symbolic identification, and symbolic comparison, and it indirectly predicted most mathematical outcomes.

Interestingly, once executive function was included as a predictor of quantitative skills, subitizing no longer accounted for any unique variance in counting, symbolic identification or comparison. Thus, when predicting other quantitative skills, it was difficult to separate the hypothetically domain-specific skill of subitizing from domain-general cognitive functioning. Note that subitizing was still a unique predictor of arithmetic fluency. The overlap between subitizing and executive functioning is not surprising if origins of the subitizing system are considered. Subitizing—the ability to quickly and exactly apprehend small quantities—has been linked to the activity of a visual-attentional system that can track up to three or four objects (Trick & Pylyshyn, 1994). On this view, individual differences in subitizing are best conceptualized as reflecting an attentional mechanism that is a general property of the cognitive system. Research on visual-spatial working memory similarly has shown a limit of three or four items. Hence, although subitizing may support the development of the counting system and thus appear to have domain-specific functions, it may be difficult to dissociate from domain-general cognitive functioning. It is also possible that the subitizing system starts off as a domain-general cognitive ability, but becomes more specialized to quantity processing across development and mathematical instruction. Disentangling these possibilities requires additional research.

Szűcs et al. (2014) failed to find any evidence that domain-specific quantitative abilities account for unique variability in math outcomes once working memory and executive function were controlled. Based on these findings, they argue for a domain-general model of mathematical skill development. However, one implication of this proposal is that improving children's domain-general skill will result in mathematical growth. It seems unlikely that children will gain mathematical skill without a basic understanding of the symbolic numerical system. It seems more reasonable to postulate that these domain-general skills assist in the development of domain-specific skills, a situation which would predict considerable overlap in domain-specific and domain-general skills. Consistent with this view, a recent large-scale longitudinal study found that domain-general executive function skills in first grade were not predictive of mathematics performance at age 15 (Watts et al., 2015). Based on these results, Watts et al. argued that executive function training would have little effect on later mathematics achievement. Although these findings do not speak to this issue directly, they are consistent with the view that executive function predicts domain-specific quantitative skills, but not mathematics performance directly. Theories which encompass the interactions between domain-general and domain-specific factors are required to better understand how mathematical knowledge develops and thus to provide guidance for instruction.

Experiential factors

Experiential factors were unique predictors of mathematical outcomes across all three studies. Age accounted for unique variance in arithmetic fluency in Studies 1 and 3, and in speeded addition and calculation knowledge in Study 3. This effect was generally small (correlation for age and arithmetic fluency was .19 in Study 1, .21 in Study 2, and .21 to .24 with arithmetic fluency factor indicators in Study 3). The effect of age is believed to reflect cohort differences (LeFevre et al., 2014). Older adults presumably spent more time on mental arithmetic during their education, and thus are faster (and in the case of calculation knowledge, which was not speeded, better) than younger adults who presumably experienced less practice. Age may also predict performance because older adults have spent more time using arithmetic in daily life. Consistent with the view that age is associated with a calculation-based advantage, and not related to superior quantitative processing more generally, age was not correlated with number line performance (Study 3) or with quantitative skills (i.e., subitizing, counting, number comparison, or order judgments) across all three studies[8].

As expected, Asian-educated participants performed better than those educated elsewhere on outcomes requiring math computation, namely arithmetic fluency (all three studies), speeded addition (Study 3), and calculation knowledge (Study 3). The effect of education location varied by outcome, demonstrating a stronger association with arithmetic fluency than with speeded addition or calculation knowledge. As with age, this effect is thought to reflect differences in mathematics experience and practice. In Asia there is a culturally-based emphasis on mathematics and substantially more time is spent on arithmetic practice relative to other cultures (e.g., Canadian culture). Interestingly, with little exception, the Asian-educated participants did not demonstrate superior quantitative processing; that is, they did not outperform those educated-elsewhere on measures of subitizing, counting, ANS, or symbolic quantitative skills. There was a small positive correlation between education location and symbolic quantity comparison (small-medium; 1 to 9) in Study 2, $r(89) = .21$, $p < .05$. Asian-educated participants were also slower at symbolic identification (i.e., naming numerals) in Study 3. This slowing is likely caused by the process of transcoding numeral words from their first language to English (Campbell & Epp, 2004). Like age, education location did not account for variance in symbolic or nonsymbolic number line performance, supporting the view that Asian-educated participants were demonstrating an advantage specific to exact

[8] The one exception was a weak correlation between age and nonsymbolic ordering of small-medium quantities in Study 3, $r(153) = .16$, $p < .05$; this is just one of many correlations, and might be spurious.

calculation. This suggests to me that once symbolic number skills are in place, specific computational practice is required for additional gains.

The results from the Full SEM with calculation knowledge as the outcome also support this view. Because this measure included algebra and fraction arithmetic, it required more specific procedural knowledge about mathematics than arithmetic fluency or speeded addition. Compared to other outcomes, the calculation knowledge measure was, a) less correlated with quantitative skills, b) less related to executive function, and c) predicted more strongly by perceived math competence. These findings suggest that as an outcome becomes more procedurally demanding, its performance relies less on domain-general and basic quantitative skills, and more on specific learning experiences. It is also important to note that the calculation knowledge measure was not speeded, whereas the quantitative measures and arithmetic tasks were; the shared speed component likely accounts for some shared variance.

Although not the focus of the present research, perceived math competence (PMC) was measured and included as a predictor of mathematical outcomes. PMC was a significant unique predictor of arithmetic fluency (all three studies), speeded addition (Study 3) and calculation knowledge (Study 3). That is, participants who felt more competent and less anxious/avoidant towards mathematics also performed better on the calculation-based outcomes. PMC was correlated with symbolic number line performance but was not a unique predictor; it was a marginally significant negative predictor of nonsymbolic number line performance. These findings suggest that ordinal knowledge and/or estimation are less strongly linked to perceived competence than are calculation skills that rely on quantitative processes. These findings are consistent with other research that has found that math anxiety and perceptions about math competence account for unique variance in mathematical outcomes (Ashcraft, 2002; Suárez-Pellicioni et al., 2016; Watt et al., 2015).

It is interesting to examine PMC in relation to basic quantitative skills, as some researchers have argued that math anxiety develops out of poor quantitative skills, such as symbolic comparison (Dietrich et al., 2015a; Maloney, Ansari, & Fugelsang, 2011; Maloney, Risko, Ansari, & Fugelsang, 2010). In this respect, there was not much consistency across the three studies. In Study 1, PMC was correlated with ANS performance; this finding is inconsistent with a recent study by Dietrich et al. (2015a), in which the ANS was not related to mathematical skill. PMC was not significantly correlated with subitizing in Studies 1 or 2. In Study 2, PMC was correlated with nonsymbolic comparison (medium; 1 to 9 dots)—which was discussed as a measure that presumably involves several quantitative systems. In Study 3, PMC was correlated with subitizing, and with symbolic comparison and ordering. Thus, if math anxiety is associated with deficits in quantitative skills, it is not clear which quantitative skills these might be, given the present findings. More specifically, the relation between PMC and arithmetic outcomes was not mediated by quantitative skills, a finding that contradicts the predictions of Maloney and colleagues. In sum, the present research does not provide support for the view that affective factors are a consequence of limitations in basic quantitative abilities.

PMC was moderately correlated with executive function in Study 3. In the present research I argue that domain-general skills contribute to the development of the domain-specific quantitative skills. How PMC fits in, in relation to executive function, quantitative processing, and mathematical performance is worth examining in future research. Although PMC was not clearly related to quantitative processing, it did account for unique variance in mathematical outcomes (beyond quantitative skills and other experiential factors), suggesting that it was an important control factor in the current research. More research is needed to explore the potential causes of the moderate but persistent relations between mathematical performance and perceived competence (Suárez-Pellicioni et al., 2016).

Strengths and Limitations

The present research was more comprehensive than most past research in which ANS and other predictors of mathematical performance were examined. In particular, these studies were the first to simultaneously examine subitizing, counting, approximate number system, and symbolic quantitative tasks in relation to each other, and as relative contributors to a variety of mathematical outcomes. It was also comprehensive in terms of considering domain-general and experiential factors that might also account for unique variance in outcomes. By using structural equation modeling I was able to simultaneously examine direct predictors of math outcomes, as well as expected indirect relations. For example, I could examine whether or not subitizing and counting predicted symbolic identification and comparison, while simultaneously examining the unique predictors of the mathematical outcome. Generally speaking, the SEMs were a good way to conceptualize and display relations among variables given the large amount of overlapping variance among constructs.

Despite these contributions, the present research was exploratory in nature. Although the hypotheses and the creation of the models were rooted in existing theory, the Full structural equation models in Study 3 were perhaps too complex given the state of the literature and the sample size. The lack of overall model fit of the Full SEMs suggests that the models were not optimally specified—that is important constructs/factors and/or pathways were missing from the models. It is also possible that a more parsimonious model would present a better solution. However, prior to the present research, it was not clear how subitizing, counting, and ANS skills would be related, because this question had not been examined before—and thus the SEMs in Study 3 are a good starting point for designing future research. In Study 3, subitizing, counting, ANS, and symbolic skills were simultaneously entered as predictors of the mathematical outcomes in order to test their relative contributions—however, given the past research, ANS skill was not expected to predict outcomes beyond symbolic quantitative skills. Removing ANS as a predictor would be one way to simplify the models; similarly, symbolic identification did not predict anything else in the quantitative pathway and could be removed. Furthermore, given the strong correlation between symbolic comparison and symbolic ordering factors, it might have been better to combine them into one factor. I did not do that in the present research because one objective was to examine symbolic comparison and ordering as unique predictors. Symbolic ordering has only recently emerged as a potentially important domain-specific quantitative skill (e.g., Lyons et al., 2014; Lyons & Beilock, 2009, 2011). Given the proposed developmental trajectory and observed patterns of relations, subitizing and counting might be better modeled as indirect rather than direct predictors of mathematical outcomes.

The present research should also be considered exploratory in nature because it employed novel tasks, which have unknown validity and reliability. For example, the subitizing measure employed in the present research may not uniquely tap subitizing skill, but may also involve domain-general skills, such as inhibitory control. When participants name a set of dots presented among other sets of dots, they may need to use inhibitory control to suppress all of the other subitizing stimuli visible to them. However, the analogous naming tasks (e.g., letter naming in Study 1 and digit naming in Study 2) were included to control for such domain-general effects. Indeed, the present evidence suggested that there was something specific about the subitizing measure (beyond sequential labeling or inhibitory control) that captured unique variance in arithmetic fluency. The arithmetic fluency measure is also relatively novel, and should be cross-validated with tests of arithmetic skill; arguably there is something specific to this measure that makes it more strongly related to the subitizing measure relative to other mathematical outcomes.

The *main* limitation of the present research is that it explores relations among quantitative skills that are theorized to *develop* in sequence—however, relations among these constructs were examined simultaneously with adult samples. With adults, one can examine relations among constructs as *clues* or *artifacts* of a developmental trajectory, but these relations are not adequate or conclusive evidence. A related limitation is that the present findings provide correlational—not causal—evidence. Nevertheless, it is my view that the current structural equation models provide a framework for future developmental research in terms of important theoretical constructs, expected relations, and prospective measures. Conducting research with young children is generally more difficult and resource-intensive than conducting research with adults. For example, developmental research requires multiple permissions and ethical approvals, and testing time is more limited than with adults due to the restricted attention spans and cognitive capacities of young children. Given these challenges, it seems reasonable to spend time developing a theoretical framework with adults prior to testing it with children, although it is also important to bear in mind that the relations among variables might be different in children than in adults.

Implications & Recommendations for Future Research

The current research highlights the potential importance of the subitizing system, which has been largely ignored in comparison to the ANS (cf. Feigenson et al. 2004; Libertus, Feigenson, & Halberda, 2013). The present research provides empirical evidence that the subitizing system is more likely to be related to mathematical outcomes (especially calculation) than the ANS. Theoretically, the small *exact* number system seems like a more plausible basis for developing mappings between symbols and quantities early in development than the approximate system (e.g., Le Corre & Carey, 2007). Nevertheless, because the present research is some of the first to examine the relative contributions of subitizing, counting, and ANS skills, more research is required prior to drawing strong conclusions about the role of subitizing as a causal agent in mathematical development, or to advising educators to develop real-world applications that emphasize subitizing as a causal factor in mathematical learning.

Research exploring early subitizing skills in relation to growth in early symbolic and mathematical skill is necessary before speculating about causal mechanisms. Based on the current findings and the existing theoretical accounts, subitizing is expected to predict counting development and early symbolic number skills. Notably, however, research tasks should be more extensively examined in terms of reliability and validity, as the present research and the ANS literature demonstrate the dangers of not doing so. When possible, future research examining subitizing, counting, and ANS across development should include domain-general measures, along with domain-specific quantitative skills to further explore the relations between these constructs—particularly between subitizing and executive functions.

Although the acquisition of counting skills has been well researched and is fairly well understood, very little research has examined subitizing skill as a predictor of symbolic number system acquisition—either concurrently or as a precursor skill (cf. LeFevre et al., 2010). The observed correlation between subitizing speed and arithmetic fluency among adults lends support to the proposed importance of the subitizing system. Future longitudinal research should examine early subitizing skills in relation to the development of counting and calculation skills; such research could potentially examine the role of subitizing and counting into adulthood.

Summary/Conclusion

The goal of the current research was to examine domain-specific quantitative abilities to gain insights about the building blocks of mathematics. Different kinds of quantitative skills were identified based on past observations and theoretical accounts. Namely, subitizing, counting, approximate number system, and symbolic quantitative skills were considered. Both the subitizing and approximate number systems are presumed to be core nonsymbolic quantitative systems. Subitizing permits the exact enumeration of small quantities, whereas the ANS is associated with approximate, inexact (i.e., relative) processing of large quantities. On the other hand, counting is the learned process of enumerating sets beyond the subitizing range to determine how many. Counting is responsible for the first mappings between nonsymbolic and symbolic quantities. To clearly distinguish between the subitizing, counting, and approximate number systems, quantitative tasks were grouped into small (1 to 3), medium (5 to 9) and large (10 to 100 quantities). This grouping of tasks permitted tests of the unique contribution of subitizing, counting, and ANS with as little measurement overlap as possible.

The present research was framed by three theoretical accounts, each of which emphasize the subitizing, counting, *or* approximate number system as the key contributor to mathematical success. Compared to the ANS literature, very little research has examined subitizing and counting skills in relation to mathematics performance with adult samples—an issue addressed by the current research. Subitizing, counting, ANS, and symbolic quantitative skills were examined in relation to each other, and as relative contributors to mathematics outcomes.

ANS and subitizing skill were not related, supporting the view that these are distinct nonsymbolic quantitative systems. ANS skill did not uniquely predict mathematical outcomes requiring exact calculation, but did predict symbolic and nonsymbolic number line performance. Counting predicted symbolic quantitative skills, but not mathematical outcomes. Furthermore, counting and subitizing were correlated, but both shared a great deal of variance with executive function. Subitizing emerged as a predictor of *arithmetic fluency* across all three studies, but did not predict other mathematical outcomes. As hypothesized, symbolic quantitative skills tended to be the strongest predictor of all mathematical outcomes, except for performance on the nonsymbolic number line. Experiential factors also predicted mathematical outcomes across all three studies. These findings provide some evidence that the subitizing system scaffolds counting and symbolic quantitative skills across development, and continues to predict arithmetic fluency in adulthood. It is recommended that future research explore the role of subitizing in the development of symbolic quantitative skills, to gain understanding of this developmental trajectory.

References

Andersson, U. (2008). Working memory as a predictor of written arithmetical skills in children: The importance of central executive functions. *The British Journal of Educational Psychology*, *78*, 181–203. doi: 10.1348/000709907X209854

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews Neuroscience*, *9*, 278–291. doi: 10.1038/nrn2334

Ansari, D., & Dhital, B. (2006). Age-related changes in the activation of the intraparietal sulcus during nonsymbolic magnitude processing: An event-related functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience, 18*, 1820–1828. doi: 10.1162/jocn.2006.18.11.1820

Ashcraft, M. H. (2002). Math anxiety and its cognitive consequences. *Current Directions in Psychological Science*, *11*, 181–185. doi: 10.1111/1467-8721.00196

Ashcraft, M. H., Krause, J. A., Hopko, D. R., Berch, D. B., & Mazzocco, M. M. M. (2007). Is math anxiety a mathematical learning disability? In D. B. Berch & M. M. M. Mazzocco (Eds.), *Why is Math so Hard for Some Children?* (pp. 329–348). Baltimore: Paul H. Brookes.

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*, 699–713. doi: 10.1037/0022-0663.96.4.699

Baddeley, A. D. (2001). Is working memory still working? *The American Psychologist*, *56*(11), 851–64. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20378789

Bartelet, D., Ansari, D., Vaessen, A., & Blomert, L. (2014). Cognitive subtypes of mathematics learning difficulties in primary education. *Research in Developmental Disabilities*, *35*, 657–670. doi: 10.1016/j.ridd.2013.12.010

Berch, D. B., Krikorian, R., & Huha, E. M. (1998). The Corsi block-tapping task: Methodological and theoretical considerations. *Brain and Cognition*, *38*, 317–338. http://doi.org/10.1006/brcg.1998.1039

Bourassa, A. (2014). *Numerical sequence recognition: Is familiarity or ordinality the primary factor in performance?* (Unpublished Master of Arts Thesis). Carleton University, Ottawa, ON, Canada.

Bouskill, C. C. (2013). *Eye movement differences between retrieval and non-retrieval strategies in simple subtraction*. (Unpublished independent research project). Carleton University, Ottawa, ON, Canada.

Brankaer, C., Ghesquière, P., & De Smedt, B. (2014). Children's mapping between non-symbolic and symbolic numerical magnitudes and its association with timed and untimed tests of mathematics achievement. *PloS One*, *9*, e93565. http://doi.org/10.1371/journal.pone.0093565

Brannon, E. M. (2002). The development of ordinal numerical knowledge in infancy. *Cognition*, *83*(3), 223–240. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11934402

Brannon, E. M. (2005). What animals know about numbers. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 85–107). New York, NY: Psychology Press.

Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science*, *282*(5389), 746–749. http://doi.org/10.1126/science.282.5389.746

Bryant, P. E., Maclean, M., Bradley, L. L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. *Developmental Psychology*, *26*(3), 429–438. http://dx.doi.org/10.1037/0012-1649.26.3.429

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*, 205–228. http://doi.org/10.1080/87565640801982312

Bull, R., & Johnston, R. S. (1997). Children's arithmetical difficulties: Contributions from processing speed, item identification and short-term memory. *Journal of Experimental Child Psychology*, *65*, 1–24. http://doi.org/0022-0965/97

Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry*, *46*(1), 3–18. http://doi.org/10.1111/j.1469-7610.2005.00374.x

Butterworth, B. (2010). Foundational numerical capacities and the origins of dyscalculia. *Trends in Cognitive Sciences*, *14*(12), 534–541. http://doi.org/10.1016/j.tics.2010.09.007

Bynner, J., & Parsons, S. (1997). *Does numeracy matter? Evidence from the National Child Development Study on the impact of poor numeracy on adult life*. London, UK: The Basics Skills Agency.

Campbell, J. I. D., & Epp, L. J. (2004). An encoding-complex approach to numerical cognition in Chinese-English bilinguals. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *58*(4), 229–244. http://doi.org/10.1037/h0087447

Carey, S. (2001). Cognitive foundations of arithmetic: Evolution and ontogenisis. *Mind and Language*, *16*(1), 37–55. http://doi.org/10.1111/1468-0017.00155

Carey, S. (2004). Bootstrapping and the origins of concepts. *Daedalus*, *133*(1), 59–68. http://doi.org/doi:10.1162/001152604772746701

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

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica*, *148*, 163–72. http://doi.org/10.1016/j.actpsy.2014.01.016

Chong, S. L., & Siegel, L. S. (2008). Stability of computational deficits in math learning disability from second through fifth grades. *Developmental Neuropsychology*, *33*, 300–317. http://doi.org/10.1080/87565640801982387

Clark, C. A. C., Nelson, J. M., Garza, J., Sheffield, T. D., Wiebe, S. A., & Espy, K. A. (2014). Gaining control: Changing relations between executive control and processing speed and their relevance for mathematics achievement over course of the preschool period. *Frontiers in Psychology*, *5*, 1-15. http://doi.org/10.3389/fpsyg.2014.00107

Clark, C. A. C., Sheffield, T. D., Wiebe, S. A., & Espy, K. A. (2013). Longitudinal associations between executive control and developing mathematical competence in preschool boys and girls. *Child Development*, *84*(2), 662–677. http://doi.org/10.1111/j.1467-8624.2012.01854.x

Clements, D. H. (1999). Subitizing: What is it ? Why teach it? *Teaching Children Mathematics*, (March), 400–405. Retrieved from http://gse.buffalo.edu/fas/clements/files/Subitizing.pdf

Cordes, S., & Brannon, E. M. (2009). Crossing the divide: Infants discriminate small from large numerosities. *Developmental Psychology*, *45*(6), 1583–94. http://doi.org/10.1037/a0015666

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review*, *8*(4), 698–707. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11848588

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, *3*(2), 63–68. http://doi.org/10.1016/j.tine.2013.12.001

Curtis, E. T. (2012). *Switching between simple addition and multiplication: Asymmetrical switch costs due to problem difficulty*. (Unpublished Master of Cognitive Science thesis). Carleton University, Ottawa, ON, Canada.

Cutini, S., Scatturin, P., Basso Moro, S., & Zorzi, M. (2014). Are the neural correlates of subitizing and estimation dissociable? An fNIRS investigation. *NeuroImage*, *85*, 391–399. http://doi.org/10.1016/j.neuroimage.2013.08.027

De Smedt, B., Noël, M.-P., Gilmore, C. K., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*, 48–55. http://doi.org/10.1016/j.tine.2013.06.001

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

Dehaene, S. (1997). *The Number Sense*. New York, NY: Oxford University Press, Inc.

Dehaene, S., & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *21*(2), 314–326. http://doi.org/10.1037/0278-7393.21.2.314

Dehaene, S., & Changeux, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience*, *5*, 390–407. http://doi.org/10.1162/jocn.1993.5.4.390

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*, 219–250. http://doi:10.1016/S0010-9452(08)70002-9

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, *21*(8), 355–61. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9720604

DeStefano, D., & LeFevre, J.-A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology*, *16*, 353–386. http://doi.org/10.1080/09541440244000328

Dietrich, J. F., Huber, S., Moeller, K., & Klein, E. (2015a). The influence of math anxiety on symbolic and non-symbolic magnitude processing. *Frontiers in Psychology*, *6*, 1–10. http://doi.org/10.3389/fpsyg.2015.01621

Dietrich, J. F., Huber, S., & Nuerk, H.-C. (2015b). Methodological aspects to be considered when measuring the approximate number system (ANS): A research review. *Frontiers in Psychology*, *6*, 1–14. http://doi.org/10.3389/fpsyg.2015.00295

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., … Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428–1246. http://doi.org/10.1037/0012-1649.43.6.1428

Fazio, L. K., Bailey, D. H., Thompson, C. a, & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, *123*, 53–72. http://doi.org/10.1016/j.jecp.2014.01.013

Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science*, *6*, 568–584. http://doi.org/10.1111/1467-7687.00313

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition*, *97*(3), 295–313. http://doi.org/10.1016/j.cognition.2004.09.010

Feigenson, L., Carey, S., & Hauser, M. D. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science*, *13*, 150–156. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11933999

Feigenson, L., Carey, S., & Spelke, E. S. (2002). Infants' discrimination of number vs. continuous extent. *Cognitive Psychology*, *44*(1), 33–66. http://doi.org/10.1006/cogp.2001.0760

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

Friso-van den Bos, I., Kroesbergen, E. H., & van Luit, J. E. H. (2014). Number sense in kindergarten children: Factor structure and working memory predictors. *Learning and Individual Differences*, *33*, 23–29. http://doi.org/10.1016/j.lindif.2014.05.003

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*, 136–148. http://doi.org/10.1111/desc.12013

Gallistel, C. R., & Gelman, R. (1991). Subitizing: The preverbal counting process. In F. Craik, W. Kessen, & A. Ortony (Eds.), *Essays in honor of George Mandler* (pp. 65–81). Hillsdale, NJ: Erlbaum.

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

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, *47*, 1539–1552. http://doi.org/10.1037/a0025510

Geary, D. C., Bow-Thomas, C. C., & Yao, Y. (1992). Counting knowledge and skill in cognitive addition: A comparison of normal and mathematically disabled children. *Journal of Experimental Child Psychology*, *54*, 372–391. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1453139

Geary, D. C., Hamson, C. O., & Hoard, M. K. (2000). Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology*, *77*, 236–263. http://doi.org/10.1006/jecp.2000.2561

Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology*, *88*, 121–151. http://doi.org/10.1016/j.jecp.2004.03.002

Geary, D. C., Hoard, M. K., & Hamson, C. O. (1999). Numerical and arithmetical cognition: Patterns of functions and deficits in children at risk for a mathematical disability. *Journal of Experimental Child Psychology*, *74*, 213–239. http://doi.org/10.1006/jecp.1999.2515

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology*, *33*, 277–299. http://doi.org/10.1080/87565640801982361

Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior Research Methods*, *43*, 981–986. http://doi.org/10.3758/s13428-011-0097-5

Gebuis, T., & Reynvoet, B. (2012). The role of visual information in numerosity estimation. *PloS One*, *7*(5), e37426. http://doi.org/10.1371/journal.pone.0037426

Gelman, R., & Butterworth, B. (2005). Number and language: How are they related? *Trends in Cognitive Sciences*, *9*, 6–10. http://doi.org/10.1016/j.tics.2004.11.004

Gelman, R., & Gallistel, C. R. (1978). *The Child's Understanding of Number*. Cambridge, Mass.: Harvard University Press.

Gelman, R., & Gallistel, C. R. (2004). Language and the origin of numerical concepts. *Science*, *306*(5695), 441–443. http://doi.org/10.1126/science.1105144

Gerstadt, L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 ½-7 years old on a Stroop-like day-night test. *Cognition*, *53*, 129–153. http://doi:10.1016/0010-0277(94)90068-X

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities*, *38*, 293–304. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16122059

Gilmore, C., Attridge, N., De Smedt, B., & Inglis, M. (2014). Measuring the approximate number system in children: Exploring the relationships among different tasks. *Learning and Individual Differences*, *29*, 50–58. http://doi.org/10.1016/j.lindif.2013.10.004

Gilmore, C. K., 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. http://doi.org/10.1371/journal.pone.0067374

Gilmore, C. K., Attridge, N., & Inglis, M. (2011). Measuring the approximate number system. *Quarterly Journal of Experimental Psychology*, *64*(11), 2099–109. http://doi.org/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. http://doi.org/10.1016/j.cognition.2010.02.002

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*, *25*, 789–798. http://doi.org/10.1177/0956797613516471

Goodwin, L. D., & Leech, N. L. (2006). Understanding correlation: Factors that affect the size of *r*. *The Journal of Experimental Education*, *74*, 249–266. doi: 10.2307/20157427

Gray, S. A., & Reeve, R. A. (2014). Preschoolers' dot enumeration abilities are markers of their arithmetic competence. *PloS One*, *9*(4), e94428. http://doi.org/10.1371/journal.pone.0094428

Gray, S. A., & Reeve, R. A. (2016). Number-specific and general cognitive markers of preschoolers' math ability profiles. *Journal of Experimental Child Psychology*, *147*, 1–21. http://doi.org/10.1016/j.jecp.2016.02.004

Gu, F. (2015). *It is all about adding*. (Unpublished BA Honours thesis). Carleton University, Ottawa, ON, Canada.

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), 1457–65. http://doi.org/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. http://doi.org/10.1038/nature07246

Hannula, M. M., Rasanen, P., & Lehtinen, E. (2007). Development of counting skills: role of spontaneous focusing on numerosity and subitizing-based enumeration. *Mathematical Thinking and Learning*, *9*(1), 51–57. http://doi.org/10.1207/s15327833mtl0901_4

Hannula-Sormunen, M. M., Lehtinen, E., & Räsänen, P. (2015). Preschool children's spontaneous focusing on numerosity, subitizing, and counting skills as predictors of their mathematical performance seven years later at school. *Mathematical Thinking and Learning*, *17*(2-3), 155–177. http://doi.org/10.1080/10986065.2015.1016814

Hauser, M. D., Carey, S., & Hauser, L. B. (2000). Spontaneous number representation in semi-free-ranging rhesus monkeys. *Proceedings. Biological Sciences / The Royal Society*, *267*(1445), 829–33. http://doi.org/10.1098/rspb.2000.1078

Hauser, M. D., Dehaene, S., Dehaene-Lambertz, G., & Patalano, A. L. (2002). Spontaneous number discrimination of multi-format auditory stimuli in cotton-top tamarins (Saguinus oedipus). *Cognition*, *86*, B23–32. http://doi.org/10.1016/S0010-0277(02)00158-0

Hecht, S. A. (2002). Counting on working memory in simple arithmetic when counting is used for problem solving. *Memory & Cognition*, *30*, 447–455. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12061765

Hecht, S., 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*, 192–227. http://doi.org/10.1006/jecp.2000.2586

Hoard, M. K., Geary, D. C., Byrd-Craven, J., & Nugent, L. (2008). Mathematical cognition in intellectually precocious first graders. *Developmental Neuropsychology*, *33*, 251–276. http://doi.org/10.1080/87565640801982338

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. http://doi.org/10.1016/j.jecp.2008.04.001

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*, 1–55. http://doi.org/10.1080/10705519909540118

Hyde, D. C. (2011). Two systems of non-symbolic numerical cognition. *Frontiers in Human Neuroscience*, *5*, 1–8. http://doi.org/10.3389/fnhum.2011.00150

Hyde, D. C., & Spelke, E. S. (2009). All numbers are not equal: An electrophysiological investigation of small and large number representations. *Journal of Cognitive Neuroscience*, *21*(6), 1039–53. http://doi.org/10.1162/jocn.2009.21090

Inglis, M., & Gilmore, C. K. (2014). Indexing the approximate number system. *Acta Psychologica*, *145*, 147–155. http://doi.org/10.1016/j.actpsy.2013.11.009

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*(3), 1221–47. http://doi.org/10.1016/j.cognition.2007.06.004

Izard, V., Pica, P., Spelke, E. S., & Dehaene, S. (2008). Exact equality and successor function: Two key concepts on the path towards understanding exact numbers. *Philosophical Psychology*, *21*(4), 491–505. http://doi.org/10.1080/09515080802285354

Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *PNAS, 106*(25), 10382–10385. http://doi.org/10.1073/pnas.0812142106

Izard, V., Streri, A., & Spelke, E. S. (2014). Toward exact number: Young children use one-to-one correspondence to measure set identity but not numerical equality. *Cognitive Psychology*, *72*, 27–53. http://doi.org/10.1016/j.cogpsych.2014.01.004

Jastak Associates. (1993). *Wide Range Achievement Test*, Rev. 3 (WRAT-3). Wilmington, DE: Wide Range Inc.

Johansson, B. S. (2005). Number-word sequence skill and arithmetic performance. *Scandinavian Journal of Psychology*, *46*(2), 157–167. http://doi.org/10.1111/j.1467-9450.2005.00445.x

Jordan, K. E., Suanda, S. H., & Brannon, E. M. (2008). Intersensory redundancy accelerates preverbal numerical competence. *Cognition*, *108*(1), 210–21. http://doi.org/10.1016/j.cognition.2007.12.001

Kamawar, D., LeFevre, J.-A., Bisanz, J., Fast, L., Skwarchuk, S.-L., Smith-Chant, B. L., & Penner-Wilger, M. (2010). Knowledge of counting principles: How relevant is order irrelevance? *Journal of Experimental Child Psychology*, *105*, 138–145. http://doi.org/10.1016/j.jecp.2009.08.004

Kleemans, T., Segers, E., & Verhoeven, L. (2011). Cognitive and linguistic precursors to numeracy in kindergarten: Evidence from first and second language learners. *Learning and Individual Differences*, *21*, 555–561. http://doi.org/10.1016/j.lindif.2011.07.008

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. http://doi.org/10.1016/j.learninstruc.2012.12.001

Koontz, K. L., & Berch, D. B. (1996). Identifying simple numerical stimuli: Processing inefficiencies exhibited by arithmetic learning disabled children. *Mathematical Cognition*, *2*(1), 1–23. http://doi.org/10.1080/135467996387525.

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*, 516–31. http://doi.org/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*, 99–125. http://doi.org/10.1016/j.cognition.2003.11.004

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*(2), 395–438. http://doi.org/10.1016/j.cognition.2006.10.005

Le Corre, M., & Carey, S. (2008). Why the verbal counting principles are constructed out of representations of small sets of individuals: A reply to Gallistel. *Cognition*, *107*(2), 650–62. http://doi.org/10.1016/j.cognition.2007.09.008

Le Corre, M., Van de Walle, G., Brannon, E. M., & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cognitive Psychology*, *52*, 130–169. http://doi.org/10.1016/j.cogpsych.2005.07.002

LeFevre, J.-A., Berrigan, L., Vendetti, C., Kamawar, D., Bisanz, J., Skwarchuk, S.-L., & Smith-Chant, B. L. (2013). The role of executive attention in the acquisition of mathematical skills for children in Grades 2 through 4. *Journal of Experimental Child Psychology*, *114*, 243–261. http://doi.org/10.1016/j.jecp.2012.10.005

LeFevre, J., Bisanz, J., & Mrkonjic, L. (1988). Cognitive arithmetic: Evidence for obligatory activation of arithmetic facts. *Memory and Cognition. 16*, 45-53. http://doi.org/ 10.3758/BF03197744

LeFevre, J.-A., DeStefano, D., Coleman, B., & Shanahan, T. (2005). Mathematical cognition and working memory. In J. I. D. Campbell (Ed.), *The handbook of mathematical cognition* (p. 361−378). New York, NY: Psychology Press.

LeFevre, J.-A., Fast, L., Skwarchuk, S.-L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development*, *81*, 1753–1767. http://doi.org/10.1111/j.1467-8624.2010.01508.x

LeFevre, J., & Liu, J. (1997) Numerical cognition: Single-digit multiplication skills of adults from China & Canada. *Mathematical Cognition, 3*, 31-62.

LeFevre, J.-A., Penner-Wilger, M., Pyke, A. A., Shanahan, T., & Deslauriers, W. A. (2014). *Putting two and two together: Declines in arithmetic fluency among young canadian adults, 1993 to 2005*. Carleton University, Ottawa, ON. Retrieved from http://www.carleton.ca/ics/research/technical-reports/view-reports/

LeFevre, J.-A., Smith-Chant, B. L., Fast, L., Skwarchuk, S.-L., Sargla, E., Arnup, J. S., … Kamawar, D. (2006). What counts as knowing? The development of conceptual and procedural knowledge of counting from kindergarten through Grade 2. *Journal of Experimental Child Psychology*, *93*(4), 285–303. http://doi.org/10.1016/j.jecp.2005.11.002

Leibovich, T., & Ansari, D. (2016). The Symbol-Grounding Problem in numerical cognition: A review of theory, evidence and outstanding questions. *Canadian Journal of Experimental Psychology*, *70*(1), 12–23. http://doi.org/10.1037/cep0000070

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*, *14*, 1292–1300. http://doi.org/10.1111/j.1467-7687.2011.01080.x

Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. *Journal of Experimental Child Psychology*, *116*(4), 829–838. http://doi.org/10.1016/j.jecp.2013.08.003

Libertus, M. E., Woldorff, M. G., & Brannon, E. M. (2007). Electrophysiological evidence for notation independence in numerical processing. *Behavioral and Brain Functions, 3*:1. (no pagination). http://doi.org/10.1186/1744-9081-3-1

Lindskog, M., Winman, A., Juslin, P., & Poom, L. (2013). Measuring acuity of the approximate number system reliably and validly: The evaluation of an adaptive test procedure. *Frontiers in Psychology*, *4*(510), 1–14. http://doi.org/10.3389/fpsyg.2013.00510

Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense: Large-number discrimination in human infants. *Psychological Science*, *14*(5), 396–401. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12930467

Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities*, *41*, 451–459. http://doi.org/10.1177/0022219408321126

Lonigan, C. J., Burgess, S. R., Anthony, J. L., & Barker, T. A. (1998). Development of phonological sensitivity in 2-to 5-year-old children. *Journal of Educational Psychology*, *90*(2), 294–311. http://doi.org/10.1037/0022-0663.90.2.294

Lyons, I. M., & Ansari, D. (2015). Foundations of children's numerical and mathematical skills: The roles of symbolic and nonsymbolic representations of numerical magnitude. *Advances in Child Development and Behavior*, *48*, 93–116. http://doi.org/10.1016/bs.acdb.2014.11.003

Lyons, I. M., Ansari, D., & Beilock, S. L. (2012). Symbolic estrangement: Evidence against a strong association between numerical symbols and the quantities they represent. *Journal of Experimental Psychology: General*, *141*, 635–641. http://doi.org/10.1037/a0027248

Lyons, I. M., Ansari, D., & Beilock, S. L. (2015). Qualitatively different coding of symbolic and nonsymbolic numbers in the human brain. *Human Brain Mapping*, *36*(2), 475–488. http://doi.org/10.1002/hbm.22641

Lyons, I. M., & Beilock, S. L. (2009). Beyond quantity: Individual differences in working memory and the ordinal understanding of numerical symbols. *Cognition*, *113*(2), 189–204. http://doi.org/10.1016/j.cognition.2009.08.003

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition*, *121*(2), 256–261. http://doi.org/10.1016/j.cognition.2011.07.009

Lyons, I. M., Nuerk, H.-C., & Ansari, D. (2015). Rethinking the implications of numerical ratio effects for understanding the development of representational precision and numerical processing across formats. *Journal of Experimental Psychology: General, 144*, 1021–1035. http://doi.org/http://dx.doi.org/10.1037/xge0000094

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1-6. *Developmental Science*, 1–13. http://doi.org/10.1111/desc.12152

Maloney, E. A., Ansari, D., & Fugelsang, J. A. (2011). The effect of mathematics anxiety on the processing of numerical magnitude. *Quarterly Journal of Experimental Psychology*, *64*(1), 10–16. http://doi.org/10.1080/17470218.2010.533278

Maloney, E. A., & Beilock, S. L. (2012). Math anxiety: Who has it, why it develops, and how to guard against it. *Trends in Cognitive Sciences, 16*(8), 404–406. http://doi.org/10.1016/j.tics.2012.06.008

Maloney, E. A., Risko, E. F., Ansari, D., & Fugelsang, J. (2010). Mathematics anxiety affects counting but not subitizing during visual enumeration. *Cognition*, *114*(2), 293–297. http://doi.org/10.1016/j.cognition.2009.09.013

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, *111*, 1–22. http://dx.doi.org/10.1037/0096-3445.111.1.1

Marzocchi, G. M., Lucangeli, D., De Meo, T., Fini, F., & Cornoldi, C. (2002). The disturbing effect of irrelevant information on arithmetic problem solving in inattentive children. *Developmental Neuropsychology*, *21*, 73–92. http://dx.doi.org/10.1207/S15326942DN2101_4

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011a). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development*, *82*, 1224–1237. http://doi.org/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. http://doi.org/10.1371/journal.pone.0023749

Mazzocco, M. M. M., & Kover, S. T. (2007). A longitudinal assessment of executive function skills and their association with math performance. *Child Neuropsychology*, *13*(1), 18–45. http://doi.org/10.1080/09297040600611346

Miller, M. R., Giesbrecht, G. F., Muller, U., McInerney, R. J., & Kerns, K. A. (2012). A latent variable approach to determining the structure of executive function in preschool children. *Journal of Cognition and Development*, *13*, 395–423. http://doi.org/10.1080/15248372.2011.585478

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. http://doi.org/10.1006/cogp.1999.0734

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 International results in mathematics.* Chestnut Hill, MA: TIMSS & PIRLS. Retrieved from http://timss.bc.edu/timss2011/international-results-mathematics.html

Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology*, *103*, 490–502. http://doi.org/10.1016/j.jecp.2009.02.003

Murphy, M. M., Mazzocco, M. M. M., Hanich, L. B., & Early, M. C. (2007). Children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities*, *40*(5), 458–478. http://doi.org/10.1177/00222194070400050901

Mussolin, C., Nys, J., Content, A., & Leybaert, J. (2014). Symbolic number abilities predict later approximate number system acuity in preschool children. *PloS One*, *9*(3), e91839. http://doi.org/10.1371/journal.pone.0091839

Muthén, B.O., & Muthén, L.K. (1998–2012). *Mplus user's guide* (7th edn.). Los Angeles, CA: Muthén & Muthén.

Myers, T., & Szűcs, D. (2015). Arithmetic memory is modality specific. *Plos One*, *10*(12), e0145614. http://doi.org/10.1371/journal.pone.0145614

Nieder, A. (2009). Prefrontal cortex and the evolution of symbolic reference. *Current Opinion in Neurobiology*, *19*(1), 99–108. http://doi.org/10.1016/j.conb.2009.04.008

Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, *32*, 185–208. http://doi.org/10.1146/annurev.neuro.051508.135550

Noël, M.-P. (2009). Counting on working memory when learning to count and to add: A preschool study. *Developmental Psychology*, *45*, 163016–43. http://doi.org/10.1037/a0016224

Noël, M.-P., & Rousselle, L. (2011). Developmental changes in the profiles of dyscalculia: An explanation based on a double exact-and-approximate number representation model. *Frontiers in Human Neuroscience*, *5*, 1–4. http://doi.org/10.3389/fnhum.2011.00165

Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PloS One*, *8*(7), e67918. http://doi.org/10.1371/journal.pone.0067918

Odic, D., Le Corre, M., & Halberda, J. (2015). Children's mappings between number words and the approximate number system. *Cognition*, *138*, 102–121. http://doi.org/10.1016/j.cognition.2015.01.008

Passolunghi, M. C., & Cornoldi, C. (2008). Working memory failures in children with arithmetical difficulties. *Child Neuropsychology*, *14*, 387–400. http://doi.org/10.1080/09297040701566662

Passolunghi, M. C., Mammarella, I. C., & Altoe, G. (2008). Cognitive abilities as precursors of the early acquisition of mathematical skills during first through second grades. *Developmental Neuropsychology*, *33*(3), 229–250. http://doi.org/10.1080/87565640801982320

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2015). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology*, (September). http://doi.org/10.1037/edu0000079

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. http://doi.org/10.1016/j.cognition.2010.03.012

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron, 53*(2), 293–305. http://doi.org/10.1016/j.neuron.2006.11.022

Piazza, M., Mechelli, A., Butterworth, B., & Price, C. J. (2002). Are subitizing and counting implemented as separate or functionally overlapping processes? *NeuroImage*, *15*(2), 435–46. http://doi.org/10.1006/nimg.2001.0980

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*(5695), 499–503. http://doi.org/10.1126/science.1102085

Price, G. R., & Fuchs, L. S. (2016). The mediating relation between symbolic and nonsymbolic foundations of math competence. *Plos One*, *11*(2), e0148981. http://doi.org/10.1371/journal.pone.0148981

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), 50–7. http://doi.org/10.1016/j.actpsy.2012.02.008

Price, J., Clement, L. M., & Wright, B. J. (2014). The role of feedback and dot presentation format in younger and older adults' number estimation. *Aging, Neuropsychology, and Cognition*, *21*(1), 68–98. http://doi.org/10.1080/13825585.2013.786015

Pusiak, R. (2015). *Arithmetic and eye-tracking: The effect of mixed and pure blocks*. (Unpublished BSc Honours thesis). Carleton University, Ottawa, ON, Canada.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, *20*, 110–122. http://doi.org/10.1016/j.lindif.2009.10.005

Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology*, *91*(2), 137–157. http://doi.org/10.1016/j.jecp.2005.01.004

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. http://doi.org/10.1037/a0027520

Reigosa-Crespo, V., González-Alemañy, E., León, T., Torres, R., Mosquera, R., & Valdés-Sosa, M. (2013). Numerical capacities as domain-specific predictors beyond early mathematics learning: A longitudinal study. *PloS One*, *8*(11), e79711. http://doi.org/10.1371/journal.pone.0079711

Reigosa-Crespo, V., Valdés-Sosa, M., Butterworth, B., Estévez, N., Rodríguez, M., Santos, E., … Lage, A. (2011). Basic numerical capacities and prevalence of developmental dyscalculia: The Havana survey. *Developmental Psychology*, *48*, 123–135. http://doi.org/10.1037/a0025356

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science*, *19*(6), 607–614. http://doi.org/10.1111/j.1467-9280.2008.02130.x

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8759042

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, *108*(3), 662–674. http://doi.org/10.1016/j.cognition.2008.05.007

Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *The British Journal of Developmental Psychology*, *30*, 344–57. http://doi.org/10.1111/j.2044-835X.2011.02048.x

Sasanguie, D., Defever, E., Maertens, B., & Reynvoet, B. (2014). The approximate number system is not predictive for symbolic number processing in kindergarteners. *Quarterly Journal of Experimental Psychology*, *67*(2), 271–280. http://doi.org/10.1080/17470218.2013.803581

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. http://doi.org/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*, 418–431. http://doi.org/10.1016/j.jecp.2012.10.012

Schildknecht, M. A. (2012). *Solving subtraction problems: Operand order affects response time and accuracy on typical problems*. (Unpublished independent research project). Carleton University, Ottawa, ON, Canada.

Schleifer, P., & Landerl, K. (2010). Subitizing and counting in typical and atypical development. *Developmental Science*, *14*, 280–291. http://doi.org/10.1111/j.1467-7687.2010.00976.x

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., & De Smedt, B. (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*, 1–16. http://doi.org/10.1111/desc.12372

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, *99*(6), 323–338. http://doi.org/10.3200/JOER.99.6.323-338

Sénéchal, M. (2006). Testing the home literacy model: Parent involvement in kindergarten is differentially related to Grade 4 reading comprehension, fluency, spelling, and reading for pleasure. *Scientific Studies of Reading*, *10*, 59–87. http://doi.org/10.1207/s1532799xssr1001_4

Sénéchal, M., & LeFevre, J.-A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development*, *73*(2), 445–460. http://doi.org/10.1111/1467-8624.00417

Sénéchal, M., Pagan, S., Lever, R., & Ouellette, G. P. (2008). Relations among the frequency of shared reading and 4-year-old children's vocabulary, morphological and syntax comprehension, and narrative skills. *Early Education & Development*, *19*(1), 27–44. http://doi.org/10.1080/10409280701838710

Sénéchal, M., & Young, L. (2008). The effect of family literacy interventions on children's acquisition of reading from Kindergarten to Grade 3: A meta-analytic review. *Review of Educational Research*, *78*(4), 880–907. http://doi.org/10.3102/0034654308320319

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*, 237–243. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12741747

Skwarchuk, S.-L., Sowinski, C., & LeFevre, J.-A. (2014). Formal and informal home learning activities in relation to children's early numeracy and literacy skills: The development of a home numeracy model. *Journal of Experimental Child Psychology*, *121*, 63–84. http://doi.org/10.1016/j.jecp.2013.11.006

Smets, K., Sasanguie, D., Szűcs, D., & Reynvoet, B. (2015). The effect of different methods to construct non-symbolic stimuli in numerosity estimation and comparison. *Journal of Cognitive Psychology*, *27*(3), 310–325. http://doi.org/10.1080/20445911.2014.996568

Soto-Calvo, E., Simmons, F. R., Willis, C., & Adams, A.-M. (2015). Identifying the cognitive predictors of early counting and calculation skills: Evidence from a longitudinal study. *Journal of Experimental Child Psychology*, *140*, 16–37. http://doi.org/10.1016/j.jecp.2015.06.011

Sowinski, C., Dunbar, K., & LeFevre, J. (2014). *Calculation Fluency Test.* (Unpublished technical report). Math Lab, Carleton University, Ottawa, Canada.

Sowinski, C., LeFevre, J.-A., Skwarchuk, S. L., Kamawar, D., Bisanz, J., & Smith-Chant, B. (2015). Refining the quantitative pathway of the Pathways to Mathematics model. *Journal of Experimental Child Psychology*, *131*, 73–93. http://doi.org/10.1016/j.jecp.2014.11.004

Starkey, P., & Cooper, R. G. (1980). Perception of numbers by human infants. *Science*, *210*, 1033–1035. Retrieved from http://faculty.psy.ohio-state.edu/opfer/lab/courses/846-Concepts_files/StarkeyCooper(1980).pdf

Steiner, E. T., & Ashcraft, M. H. (2012). Three brief assessments of math achievement. *Behavior Research Methods*. http://doi.org/10.3758/s13428-011-0185-6

Stock, P., Desoete, A., & Roeyers, H. (2009). Predicting arithmetic abilities: The role of preparatory arithmetic markers and intelligence. *Journal of Psychoeducational Assessment*, *27*, 237–251. http://doi.org/10.1177/073482908330587

Suárez-Pellicioni, M., Núñez-Peña, M. I., & Colomé, À. (2016). Math anxiety: A review of its cognitive consequences, psychophysiological correlates, and brain bases. *Cognitive, Affective & Behavioural Neuroscience, 16*, 3–22. http://doi.org/10.3758/s13415-015-0370-7

Swanson, H. L., & Kim, K. (2007). Working memory, short-term memory, and naming speed as predictors of children's mathematical performance. *Intelligence*, *35*(2), 151–168. http://doi.org/10.1016/j.intell.2006.07.001

Szűcs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2014). Cognitive components of a mathematical processing network in 9-year-old children. *Developmental Science*, *17*(4), 506–524. http://doi.org/10.1111/desc.12144

Toll, S. W. M., & Van Luit, J. E. H. (2013). The development of early numeracy ability in kindergartners with limited working memory skills. *Learning and Individual Differences*, *25*, 45–54. http://doi.org/10.1016/j.lindif.2013.03.006

Torbeyns, J., Verschaffel, L., & Ghesquière, P. (2004). Strategic aspects of simple addition and subtraction: The influence of mathematical ability. *Learning and Instruction*, *14*, 177–195. http://doi.org/10.1016/j.learninstruc.2004.01.003

Treutlein, A., Zöller, I., Roos, J., & Schöler, H. (2009). Effects of phonological awareness training on reading achievement. *Written Language & Literacy*, *11*(2), 147–166. http://doi.org/10.1075/wll.11.2.03tre

Trick, L. M. (2008). More than superstition: Differential effects of featural heterogeneity and change on subitizing and counting. *Perception and Psychophysics*, *70*(5), 743–760. http://doi.org/10.3758/PP.70.5.743

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, *101*(1), 80–102. http://dx.doi.org/10.1037/0033-295X.101.1.80

Vanbinst, K., Ansari, D., Ghesquière, P., & De Smedt, B. (2016). Symbolic numerical magnitude processing is as important to arithmetic as phonological awareness is to reading. *PloS One*, *11*(3), e0151045. http://doi.org/10.1371/journal.pone.0151045 vanMarle, K., Chu, F. W., Li, Y., & Geary, D. C. (2014). Acuity of the approximate number system and preschoolers' quantitative development. *Developmental Science*, *17*, 492–505. http://doi.org/10.1111/desc.12143

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its casual role in the acquisition of reading skills. *Psychological Bulletin*, *101*(2), 192–212. http://dx.doi.org/10.1037/0033-2909.101.2.192

Watts, T. W., Duncan, G. J., Chen, M., Claessens, A., Davis-Kean, P. E., Duckworth, K., … Susperreguy, M. I. (2015). The role of mediators in the development of longitudinal mathematics achievement associations. *Child Development*, *86*(6), 1892–1907. http://doi.org/10.1111/cdev.12416

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence™ - Second Edition (WASI™-II)*. San Antonio, TX: Pearson Assessments.

Wu, S. S., Meyer, M. L., Maeda, U., Salimpoor, V., Tomiyama, S., Geary, D. C., & Menon, V. (2008). Standardized assessment of strategy use and working memory in early mental arithmetic performance. *Developmental Neuropsychology*, *33*, 365–393. http://doi.org/10.1080/87565640801982445

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*, 155–193. http://doi:10.1016/0010-0277(90)90003-3

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11. http://doi:10.1016/S0010-0277(99)00066-9

Appendices

Appendix A: Study 2 Rapid Comparison Stimuli

Section	Left	Right	Correct	Proportion	Control Type
Sample Items	6	3	Left	0.50	PC
	8	2	Left	0.25	AC
	3	9	Right	0.33	AC
	6	9	Right	0.67	PC
Practice Items	9	3	Left	0.33	AC
	2	4	Right	0.50	PC
	3	2	Left	0.67	PC
	6	2	Left	0.33	AC
	3	6	Right	0.50	AC
	8	4	Left	0.50	PC
	2	8	Right	0.25	PC
	2	6	Right	0.33	AC
Form A	8	6	Left	0.75	PC
	2	7	Right	0.29	AC
	4	9	Right	0.44	PC
	9	5	Left	0.56	AC
	1	8	Right	0.13	AC
	6	5	Left	0.83	AC
	3	8	Right	0.38	PC
	8	7	Left	0.88	PC
	4	1	Left	0.25	AC
	9	2	Left	0.22	PC
	1	6	Right	0.17	PC
	3	5	Right	0.60	AC
	3	1	Left	0.33	AC
	4	8	Right	0.50	AC
	6	4	Left	0.67	PC
	4	5	Right	0.80	PC
	8	5	Left	0.63	PC
	4	7	Right	0.57	PC
	9	1	Left	0.11	AC
	8	9	Right	0.89	AC
	5	2	Left	0.40	AC
	1	5	Right	0.20	PC
	6	7	Right	0.86	AC
	4	6	Right	0.67	PC
	6	8	Right	0.75	PC
	7	1	Left	0.14	PC
	7	5	Left	0.71	AC
	3	7	Right	0.43	AC
	2	1	Left	0.50	AC
	9	7	Left	0.78	PC

Appendix A: Study 2 Rapid Comparison Stimuli -continued-

Section	Left	Right	Correct	Proportion	Control Type
	7	2	Left	0.29	PC
Form B	5	6	Right	0.83	AC
	5	1	Left	0.20	AC
	1	3	Right	0.33	PC
	1	4	Right	0.25	AC
	2	5	Right	0.40	PC
	6	1	Left	0.17	PC
	7	3	Left	0.43	PC
	5	8	Right	0.63	PC
	9	4	Left	0.44	AC
	5	4	Left	0.80	PC
	3	4	Right	0.75	AC
	7	4	Left	0.57	AC
	1	9	Right	0.11	AC
	7	8	Right	0.88	PC
	9	6	Left	0.67	PC
	7	9	Right	0.78	PC
	8	1	Left	0.13	PC
	5	3	Left	0.60	AC
	2	3	Right	0.67	AC
	1	2	Right	0.50	PC
	5	9	Right	0.56	AC
	9	8	Left	0.89	AC
	7	6	Left	0.86	PC
	8	3	Left	0.38	AC
	4	2	Left	0.50	AC
	5	7	Right	0.71	AC
	4	3	Left	0.75	PC
	1	7	Right	0.14	PC
	2	9	Right	0.22	AC

Note. AC = Area correlated; numerosity and area were positive correlated.
PC = Perimeter controlled and numerosity negatively correlated with area.

Appendix B: Comparison of Rapid Identification and Arithmetic Fluency (CFT) Across Studies

	Rapid Identification						Rapid Comparison S2		Arithmetic Fluency (CFT)		
		Small Quantities									
	Letters	Symbolic		Nonsymbolic							
	S1	S2	S3	S1	S2	S3	S	NS	S1	S2	S3
Min	2.1	2.0	2.0	1.6	1.8	1.6	0.7	0.5	9	6	6
Max	5.3	5.6	4.8	4.1	4.2	3.6	2.7	2.4	98	81	84
Mean	3.3	3.3	3.3	2.7	2.7	2.6	1.5	1.2	38	32	35
SD	0.65	0.71	0.66	0.48	0.49	0.39	0.34	0.32	16.9	16.0	16.2

Note. Rapid identification tasks were items-per-second. Arithmetic fluency scores were total correct on the Calculation Fluency Test (CFT). S1 = Study 1, S2 = Study 2, S3 = Study 3.

Appendix C: Study 3 Small-Medium Comparison Stimuli

Stimuli L	Stimuli R	Proportion	Control Type	Category
1	9	0.11	PC	Mixed
9	1	0.11	AC	Mixed
1	8	0.13	PC	Mixed
8	1	0.13	AC	Mixed
1	7	0.14	AC	Mixed
7	1	0.14	PC	Mixed
1	6	0.17	n/a	Mixed[a]
6	1	0.17	n/a	Mixed[a]
1	5	0.20	AC	Mixed
5	1	0.20	PC	Mixed
2	9	0.22	AC	Mixed
9	2	0.22	PC	Mixed
1	4	0.25	AC	Mixed[b]
1	4	0.25	PC	Mixed
1	4	0.25	PC	Mixed
2	8	0.25	PC	Mixed
4	1	0.25	AC	Mixed
4	1	0.25	AC	Mixed[b]
4	1	0.25	PC	Mixed
8	2	0.25	AC	Mixed
2	7	0.29	PC	Mixed
7	2	0.29	AC	Mixed
1	3	0.33	AC	Subitizing[b]
1	3	0.33	PC	Subitizing
1	3	0.33	PC	Subitizing
2	6	0.33	AC	Mixed
3	1	0.33	AC	Subitizing
3	1	0.33	AC	Subitizingb
3	1	0.33	PC	Subitizing
3	9	0.33	AC	Mixed
6	2	0.33	PC	Mixed
9	3	0.33	PC	Mixed
3	8	0.38	PC	Mixed
8	3	0.38	AC	Mixed
2	5	0.40	AC	Mixed
5	2	0.40	PC	Mixed

Note. L=Left. R=Right. PC=Perimeter Controlled. AC=Area Controlled.
[a]Trial only appears in digit comparison and only for early participants.
[b]Trial only appears in quantity comparison.

Appendix C: Study 3 Small-Medium Comparison Stimuli -continued-

Stimuli		Proportion	Control Type	Category
L	R			
3	7	0.43	AC	Mixed
7	3	0.43	PC	Mixed
4	9	0.44	AC	Counting
9	4	0.44	PC	Counting
1	2	0.50	AC	Subitizing
1	2	0.50	AC	Subitizing[b]
1	2	0.50	PC	Subitizing
2	1	0.50	AC	Subitizing[b]
2	1	0.50	PC	Subitizing
2	1	0.50	PC	Subitizing
2	4	0.50	AC	Mixed[b]
2	4	0.50	PC	Mixed
2	4	0.50	PC	Mixed
3	6	0.50	PC	Mixed
4	2	0.50	AC	Mixed
4	2	0.50	AC	Mixed[b]
4	2	0.50	PC	Mixed
4	8	0.50	AC	Counting
6	3	0.50	AC	Mixed
8	4	0.50	PC	Counting
5	9	0.56	AC	Counting
9	5	0.56	PC	Counting
4	7	0.57	AC	Counting
7	4	0.57	PC	Counting
3	5	0.60	PC	Mixed
5	3	0.60	AC	Mixed
5	8	0.63	PC	Counting
8	5	0.63	AC	Counting
2	3	0.67	AC	Subitizing[b]
2	3	0.67	PC	Subitizing
3	2	0.67	AC	Subitizing[b]
3	2	0.67	PC	Subitizing
4	6	0.67	AC	Counting
6	4	0.67	PC	Counting
6	9	0.67	AC	Counting
9	6	0.67	PC	Counting

Note. L=Left. R=Right. PC=Perimeter Controlled. AC=Area Controlled.
[a]Trial only appears in digit comparison and only for early participants.
[b]Trial only appears in quantity comparison.

Appendix C: Study 3 Small-Medium Comparison Stimuli -continued-

Stimuli		Proportion	Control Type	Category
L	R			
5	7	0.71	PC	Counting
7	5	0.71	AC	Counting
3	4	0.75	AC	Mixed
3	4	0.75	AC	Mixed[b]
3	4	0.75	PC	Mixed
4	3	0.75	AC	Mixed[b]
4	3	0.75	PC	Mixed
5	6	0.83	PC	Counting
6	5	0.83	AC	Counting
6	7	0.86	PC	Counting
7	6	0.86	AC	Counting
7	8	0.88	PC	Counting
8	7	0.88	AC	Counting
8	9	0.89	PC	Counting
9	8	0.89	AC	Counting

Note. L=Left. R=Right. PC=Perimeter Controlled. AC=Area Controlled.
[a]Trial only appears in digit comparison and only for early participants.
[b]Trial only appears in quantity comparison.

Appendix D: Study 3 Large Comparison Stimuli

Stimuli		Proportion	Control Type
Left	Right		
11	99	0.11	RPC
99	11	0.11	LAC
14	98	0.14	RPC
98	14	0.14	LAC
13	65	0.20	RPC
18	90	0.20	RAC
65	13	0.20	LAC
90	18	0.20	LPC
14	56	0.25	RPC
18	72	0.25	RPC
20	80	0.25	RAC
56	14	0.25	LAC
72	18	0.25	LAC
80	20	0.25	LPC
13	39	0.33	RPC
19	57	0.33	RAC
23	69	0.33	RAC
29	87	0.33	RPC
57	19	0.33	LPC
69	23	0.33	LPC
87	29	0.33	LAC
14	35	0.40	RPC
30	75	0.40	RAC
34	85	0.40	RPC
35	14	0.40	LAC
75	30	0.40	LPC
85	34	0.40	LAC
16	36	0.44	RPC
28	63	0.44	RAC
36	16	0.44	LAC
40	90	0.44	RPC
63	28	0.44	LPC
90	40	0.44	LAC
15	27	0.56	RPC
20	36	0.56	RAC
27	15	0.56	LAC
36	20	0.56	LPC

Note. L=Left. R=Right.
RPC=Right Perimeter Controlled. LPC=Left Perimeter Controlled.
LAC=Left Area Controlled. RAC=Right Area Controlled.

Appendix D: Study 3 Large Comparison Stimuli -continued-

Stimuli Left	Stimuli Right	Proportion	Control Type
45	81	0.56	RPC
81	45	0.56	LAC
27	45	0.60	RAC
30	50	0.60	RPC
45	27	0.60	LPC
50	30	0.60	LAC
57	95	0.60	RAC
95	57	0.60	LPC
16	24	0.67	RPC
24	16	0.67	LAC
40	60	0.67	RAC
60	40	0.67	LPC
64	96	0.67	RPC
96	64	0.67	LAC
18	24	0.75	RPC
24	18	0.75	LAC
42	56	0.75	RAC
56	42	0.75	LPC
63	84	0.75	RPC
72	96	0.75	RAC
84	63	0.75	LAC
96	72	0.75	LPC
12	15	0.80	RPC
15	12	0.80	LAC
24	30	0.80	RAC
30	24	0.80	LPC
32	40	0.80	RAC
40	32	0.80	LPC
44	55	0.80	RPC
55	44	0.80	LAC
60	75	0.80	RPC
68	85	0.80	RAC
75	60	0.80	LAC
76	95	0.80	RAC
85	68	0.80	LPC
95	76	0.80	LPC
15	18	0.83	RPC
18	15	0.83	LAC
30	36	0.83	RAC

Note. L=Left. R=Right.
RPC=Right Perimeter Controlled. LPC=Left Perimeter Controlled.
LAC=Left Area Controlled. RAC=Right Area Controlled.

Appendix D: Study 3 Large Comparison Stimuli -continued-

Stimuli Left	Stimuli Right	Proportion	Control Type
35	42	0.83	RAC
36	30	0.83	LPC
42	35	0.83	LPC
50	60	0.83	RPC
60	50	0.83	LAC
65	78	0.83	RPC
75	90	0.83	RAC
78	65	0.83	LAC
80	96	0.83	RAC
90	75	0.83	LPC
96	80	0.83	LPC
12	14	0.86	RPC
14	12	0.86	LAC
18	21	0.86	RAC
21	18	0.86	LPC
30	35	0.86	RAC
35	30	0.86	LPC
42	49	0.86	RPC
49	42	0.86	LAC
60	70	0.86	RPC
70	60	0.86	LAC
72	84	0.86	RAC
84	72	0.86	LPC
84	98	0.86	RAC
98	84	0.86	LPC
21	24	0.88	RPC
24	21	0.88	LAC
35	40	0.88	RAC
40	35	0.88	LPC
49	56	0.88	RAC
56	49	0.88	LPC
56	64	0.88	RPC
63	72	0.88	RPC
64	56	0.88	LAC
70	80	0.88	RAC
72	63	0.88	LAC
80	70	0.88	LPC
84	96	0.88	RAC
96	84	0.88	LPC
16	18	0.89	RPC
18	16	0.89	LAC
24	27	0.89	RAC

Note. L=Left. R=Right.
RPC=Right Perimeter Controlled. LPC=Left Perimeter Controlled.
LAC=Left Area Controlled. RAC=Right Area Controlled.

Appendix D: Study 3 Large Comparison Stimuli -continued-

Stimuli Left	Stimuli Right	Proportion	Control Type
27	24	0.89	LPC
40	45	0.89	RAC
45	40	0.89	LPC
48	54	0.89	RPC
54	48	0.89	LAC
56	63	0.89	RPC
63	56	0.89	LAC
72	81	0.89	RAC
80	90	0.89	RAC
81	72	0.89	LPC
90	80	0.89	LPC
18	20	0.90	RPC
20	18	0.90	LAC
27	30	0.90	RAC
30	27	0.90	LPC
36	40	0.90	RAC
40	36	0.90	LPC
45	50	0.90	RPC
50	45	0.90	LAC
54	60	0.90	RPC
60	54	0.90	LAC
63	70	0.90	RAC
70	63	0.90	LPC
81	90	0.90	RAC
90	81	0.90	LPC
10	11	0.91	RPC
11	10	0.91	LAC
20	22	0.91	RPC
22	20	0.91	LAC
30	33	0.91	RAC
33	30	0.91	LPC
40	44	0.91	RAC
44	40	0.91	LPC
60	66	0.91	RPC
66	60	0.91	LAC
70	77	0.91	RPC
77	70	0.91	LAC
80	88	0.91	RAC
88	80	0.91	LPC
90	99	0.91	RAC
99	90	0.91	LPC
11	12	0.92	RPC

Note. L=Left. R=Right.
RPC=Right Perimeter Controlled. LPC=Left Perimeter Controlled.
LAC=Left Area Controlled. RAC=Right Area Controlled.

Appendix D: Study 3 Large Comparison Stimuli -continued-

Stimuli Left	Stimuli Right	Proportion	Control Type
12	11	0.92	LAC
22	24	0.92	RPC
24	22	0.92	LAC
33	36	0.92	RAC
36	33	0.92	LPC
44	48	0.92	RAC
48	44	0.92	LPC
55	60	0.92	RPC
60	55	0.92	LAC
66	72	0.92	RPC
72	66	0.92	LAC
77	84	0.92	RAC
84	77	0.92	LPC
88	96	0.92	RAC
96	88	0.92	LPC

Note. L=Left. R=Right.
RPC=Right Perimeter Controlled. LPC=Left Perimeter Controlled.
LAC=Left Area Controlled. RAC=Right Area Controlled.

Appendix E: Study 3 Ordering Stimuli

Study 3 small-medium range (1 - 9) symbolic and nonsymbolic ordering stimuli

In order				Not in order			
Trial	Stim 1	Stim 2	Stim 3	Trial	Stim 1	Stim 2	Stim 3
	1	2	3		2	3	1
	1	2	3		1	3	2
	2	3	4		4	2	3
	3	4	5		3	5	4
	4	5	6		6	4	5
	5	6	7		5	7	6
	6	7	8		7	6	8
	7	8	9		7	9	8
	1	3	5		3	1	5
	2	4	6		4	6	2
	3	5	7		5	7	3
	4	6	8		4	8	6
	1	4	7		1	7	4
	1	5	9		5	1	9
	2	5	8		2	8	5
	3	6	9		9	3	6
	5	7	9		5	9	7
	1	3	4		1	4	3
	2	5	7		5	7	2
	3	5	8		3	8	5
	4	5	9		4	9	5
	1	4	5		4	5	1
	1	5	6		1	6	5
	2	7	9		2	9	7
	3	4	7		3	7	4
	1	6	7		1	7	6
	1	7	8		1	8	7
	1	8	9		1	9	8
	2	6	8		2	8	6
	1	2	4		2	4	1
	1	2	8		1	8	2
	1	3	8		1	8	3
	1	4	6		4	1	6
	1	7	9		1	9	7
	2	3	8		8	2	3
	2	6	7		7	2	6
	3	4	6		4	3	6
	3	5	9		3	9	5
	4	5	7		7	4	5
	4	6	9		6	9	4
	4	7	9		4	9	7
	5	6	9		9	5	6
	6	7	9		7	6	9

Appendix E: Study 3 Ordering Stimuli

Study 3 large range (10 – 100) symbolic and nonsymbolic ordering stimuli

In order				Not in order			
Trial	Stim 1	Stim 2	Stim 3	Trial	Stim 1	Stim 2	Stim 3
	15	30	45	44.	30	45	15
	26	52	78	45.	26	78	52
	28	42	56	46.	56	28	42
	24	32	40	47.	24	40	32
	60	75	90	48.	90	60	75
	30	36	42	49.	30	42	36
	18	21	24	50.	21	18	24
	35	40	45	51.	35	45	40
	19	57	95	52.	57	19	95
	32	64	96	53.	64	96	32
	27	45	63	54.	45	63	27
	32	48	64	55.	32	64	48
	13	52	91	56.	13	91	52
	11	55	99	57.	55	11	99
	18	45	72	58.	18	72	45
	12	24	36	59.	36	12	24
	40	56	72	60.	40	72	56
	25	75	100	61.	25	100	75
	12	30	42	62.	30	42	12
	12	20	32	63.	12	32	20
	24	30	54	64.	24	54	30
	15	60	75	65.	60	75	15
	16	80	96	66.	16	96	80
	16	56	72	67.	16	72	56
	21	28	49	68.	21	49	28
	13	78	91	69.	13	91	78
	11	77	88	70.	11	88	77
	10	80	90	71.	10	90	80
	12	36	48	72.	36	48	12
	21	44	89	73.	21	89	44
	12	25	95	74.	12	95	25
	10	34	88	75.	10	88	34
	17	64	95	76.	64	17	95
	10	78	99	77.	10	99	78
	10	16	41	78.	41	10	16
	29	84	99	79.	99	29	84
	26	35	54	80.	35	26	54
	12	21	37	81.	12	37	21
	48	59	85	82.	85	59	48
	29	42	62	83.	42	62	29
	23	57	72	84.	23	72	57
	46	53	81	85.	81	46	53
	30	34	46	86.	34	46	30