11-8-2021 1:30 PM

Cognitive and Affective Correlates of Math Achievement: An Examination of the Roles of Symbolic Number Development and Math Anxiety

Nathan T.T. Lau, *The University of Western Ontario*

Supervisor: Ansari, Daniel, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Psychology

Abstract

The principal aim of this thesis was to examine the correlates of children's math achievement. Studies 1 and 2 provided insights regarding the relation between children's learning of symbolic numbers and math achievement. Study 3 examined the relation between math anxiety and math achievement.

There is currently significant debate regarding the role that the approximate, non-verbal number ability (approximate number system) plays in the development of children's understanding of symbolic numbers. In particular, there exists much discussion regarding whether children's approximate number ability predicts later symbolic number ability (the mapping account) or the other way around (the refinement account). Study 1 compared the theoretical predictions of these two developmental accounts over three timepoints in children from senior kindergarten to grade 1 (N=622). Results suggest that symbolic number ability consistency predicts later approximate number ability and math achievement, supporting the refinement account.

Emergent research findings suggest that there may be heterogeneity in the developmental trajectory children undergo when learning symbolic numbers. Study 2 examined the degree to which the theoretical predictions laid out by the mapping and refinement accounts adequately describe the whole population using the same sample as in study 1. Results suggest that children's developmental trajectories are remarkably homogenous, however, there is considerable heterogeneity regarding the speed in which children progress through levels of development. Taken together, results from studies 1 and 2 suggest that the refinement account best describes children's symbolic number development, and that a vast majority of students follow the same developmental path.

Using three international studies of student achievement, study 3 uses multilevel modelling to examine the degree to which the level of math anxiety in children's immediate cultural and educational contexts could predict their math achievement over and above their own math anxiety. Results suggest that math anxiety in the education environment predicts math achievement. However, there are significant between-country differences in the strength of this contextual effect of math anxiety. These results bring into question the generalizability of existing research findings and suggest that children’s cultural and educational contexts must be taken into account for a complete description of the relation.

Keywords

Approximate Number System, Numerical Cognition, Math Anxiety, Math Achievement, Education, Culture

Summary for Lay Audience

Math achievement in early childhood has been shown to be an important factor for success in school, career advancement, and overall well-being. Consequently, there is great interest in understanding the underlying factors that contribute to improving children's math achievement. The current thesis contributes to this goal by examining two factors that affect math achievement: 1) the process in how children initially learn numerical symbols (e.g., Arabic numbers, spoken numbers, etc.), 2) children's anxiety towards doing math problems.

Research over the past decades has revealed that, even as infants, children possess an innate cognitive system that can perceive numerical quantities (approximate number system). Using data from 622 children, study 1 investigates whether children's approximate number system predicts their understanding of numerical symbols, or is it the other way around. Results suggest that increased understanding in children's understanding of numerical symbols not only refines and improves children's approximate number system, but also predicts math achievement.

Study 2 examines the degree to which all children (or only some groups of children) undergo the developmental process as described in Study 1. Results indicate that virtually all children undergo the same developmental process, but they progress through these developmental levels at different speeds. Together, results from studies 1 and 2 suggest that recognizing and using symbolic numbers is an important skill that predicts future math achievement, and virtually all children undergo the same developmental process.

Using three international studies of achievement, study 3 asks whether anxiety towards mathematics in one's educational peer group could affect one's math achievement. In other words, this study asks whether there would be a difference in a child's math achievement if his or her classmates have high average math anxiety versus low average math anxiety. Results suggest that higher levels of peer's math anxiety are associated with a reduction in one's own math achievement. However, this effect is only observed in approximately half

the countries. This research inspires future studies to examine why there might be these between-country differences.

Together, the current thesis provides new insights into the ways different early childhood factors contribute to influencing children's math achievement.

Co-Authorship Statement

The work presented in this thesis was designed and written in collaboration with my advisor, Dr. Daniel Ansari. For each of the studies in this thesis, Dr. Ansari contributed to the design, analysis, and interpretation of the findings.

For study 1, Dr. Rebecca Merkley from Carlton University and Dr. Paul Tremblay from the University of Western Ontario contributed to the design, analysis and interpretation of the findings. Samuel Zhang and Stefanie De Jesus from the Toronto District School Board aided in data collection.

For study 2, Samuel Zhang and Stefanie De Jesus from the Toronto District School Board aided in data collection.

For study 3, Dr. Zack Hawes from the University of Toronto and Dr. Paul Tremblay from the University of Western Ontario contributed to the design, analysis and interpretation of the findings.

Acknowledgements

First, I am extremely grateful and lucky to complete my PhD in Dr. Daniel Ansari's Numerical Cognition Lab. The past four years have been nothing but positive experiences and have sharpened my critical thinking skills. Your curiosity and infectious enthusiasm towards numerical cognition is really inspiring; it is always a pleasure to talk with you on research. Your mentorship, guidance, support, and encouragement helped me immensely. I could not have imagined a better supervisor. Thank you.

I am also grateful to Dr. Paul Tremblay. It was through attending your classes in SEM and seeing the passion you showed for Structural Equation Modelling that really stirred me to dig deeper into using quantitative methods in forwarding research. I am immensely grateful for your patience in answering my questions and your aid in helping me translate theory to practice.

I am also grateful for all current and former members of the Numerical Cognition Lab: Aymee Alvarez, Pierina Cheung, Rebekka Lagace Cusiac, Bea Goffin, Celia Goffin, Zack Hawes, Rebecca Merkley, Lien Peters, Michael Slipenkyj, Mojtaba Soltanlou, Moriah Sokolowski, and Eric Wilkey. It has been such a pleasure to learn from you all.

I would also like to thank my examining committee: Dr. Drew Bailey, Dr. Mark Cleveland, Dr. Bruce Morton and Dr. John Sakaluk.

Finally, I would like to thank my family. To my mom and dad for your unconditional love and support. To my sister, for always being there for me.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

1.1 General Introduction

The impact and importance of math in our lives is significant. Like literacy, math provides a powerful tool to describe, understand, and shape the world around us. Math skills are predictive of education success, future employment, the likelihood of entering and succeeding in STEM professions (Science, Technology, Engineering, Mathematics), economic productivity, financial literacy, socioeconomic status (SES), incarceration rates, health, and overall well-being (e.g., Duncan et al., 2007; Nicholls et al., 2007; Parsons & Bynner, 2005; Ritchie & Bates, 2013; Skagerlund et al., 2018; Williams et al., 2003). The overarching goal of this thesis is to contribute to the ongoing effort in understanding the factors in childhood that predict later math achievement. To this end, three studies were carried out. The first two studies aimed to contribute to the hotly debated topic of how children initially learn symbolic numbers, and how individual differences in symbolic number learning may be related with math achievement. The third study focused on the emerging issue of math anxiety and how it relates to math achievement.

The central issue addressed in Studies 1 and 2 is concerned with the relation between individual differences in how children initially acquire symbolic number knowledge and their developing math skills. Divergent research findings in recent decades have given birth to two different and often contradictory developmental accounts of how children initially learn numerical symbols, and there is ongoing discussion regarding which of the two competing developmental accounts better describe children's numerical development. Studies 1 and 2 aimed to contribute to this ongoing discussion by examining the reasons behind this divergence in findings. In doing so, these studies intend to clarify this discrepancy, and contribute to identifying the developmental account that best approximates reality.

Study 3 approached the emerging issue of how students' math anxiety may be related with math achievement. While substantial research has been conducted in an effort to find the predictors of math anxiety and assess math anxiety's effect on students, research findings as a whole remain siloed and isolated. Consequently, this has blindsided researchers and has left multiple important research questions unanswered. Using multiple large-scale international studies of achievement, study 3 addresses three pressing research questions that advances understanding of math anxiety.

1.2 Background for Studies 1 and 2

1.2.1 Symbolic Number Learning and Math Achievement

Studies of children's development of numerical and mathematical skills have consistently revealed that there exist large individual differences in these skills from an early age (e.g, De Smedt et al., 2009; Jordan et al., 2007). Importantly, it has been demonstrated that this variability in early childhood is not only a strong predictor of children's later math achievement, but also a predictor of the rate of growth in math ability that children experience over the course of schooling (Jordan et al., 2009). Taken together, it is not surprising that children who are behind their peers in basic numerical competencies at the beginning of elementary school are at high risk of staying behind in future grades (Duncan et al., 2007). Ultimately, low math competency may cause students to shun math and tasks requiring math, which may lead to future limitations in vocational choices and selections (Nicholls et al., 2007). Given the gravity of the negative outcomes that low math performance early in childhood may catalyze, increasing research attention has been placed to better understand the various factors that predict early math achievement.

One prominent area of research relevant to early math achievement pertains to the process through which children in early childhood initially acquires the uniquely human ability to represent quantities with arbitrary (i.e., non-iconic) symbols (e.g., with Arabic numerals, spoken words, etc.; Dehaene, 1999). This ability to represent quantities with symbols (e.g., knowing that the number word "3" stands for all possible sets of three items etc.) is a foundational skill on which most, if not all contemporary mathematics is based upon. Given this foundational role, it has been hypothesized that individual differences in math achievement may be rooted in individual differences in the efficiency through which children initially learns to represent quantities with symbols (e.g., Halberda & Feigenson, 2008). Therefore, enrichment programs that aid children in acquiring symbolic number knowledge and intervention programs that could identify children who are struggling with the acquiring symbolic number knowledge may be promising avenues to bolster children's early math achievement.

However, in recent years, two different, and sometimes contradictory developmental accounts have been forwarded to explain how children initially acquire the skill to represent quantities with symbols. These developmental accounts postulate different mechanisms that underlie the learning of symbolic number representational abilities and consequently, posit different skills that should be the target of enrichment and intervention. Identifying the developmental account that best fits the developmental data is a priority, as the effectiveness of enrichment and intervention programs are contingent on the accuracy of the underlying developmental account on which they are based. To begin, we will first review these two developmental accounts and highlight the theoretical predictions that each account posits.

1.2.2 The Mapping Account

Even in infancy, humans exhibit a remarkable sensitivity to numerical information (Antell & Keating, 1983; Cheries et al., 2006; Van Loosbroek & Smitsman, 1990; Xu & Spelke, 2000). For instance, using a habituation paradigm, it has been shown that 6-month-old infants can discriminate between arrays of dots with different quantities (Xu & Spelke, 2000). Concurrently, it has been observed that multiple, diverse animal species, ranging from dolphins to chimpanzees, are sensitive to changes in numerosity (Agrillo, 2015; Nieder & Merten, 2007; Viswanathan & Nieder, 2013). Taken together, the confluence of evidence suggests that the ability to discriminate between numerosities is an innate and evolutionarily ancient cognitive ability.

This ability is often referred to as the approximate number system (ANS) because representations of the perceived quantities are characterized by the Weber-Fechner Law.

When comparing two sets of non-symbolic quantities, it becomes increasingly difficult to distinguish between the two sets as the numerical distance between the quantities grows closer (the distance effect; e.g., Cordes, Gelman, Gallistel, & Whalen, 2001), and for any particular fixed numerical distance between two sets of quantities, it becomes increasingly difficult to distinguish between the two sets when the average magnitude of the two quantities increases (the size effect; e.g., Moyer & Landauer, 1967).

According to the mapping account of development, this innate quantity representation ability forms a foundation that enables for the learning of symbolic numbers (Dehaene, 2007; Feigenson et al., 2004; Piazza, 2011). Specifically, this hypothesis posits that the learning of symbolic numbers co-opts the same neural circuitry for processing approximate magnitudes, and suggested that because of this, both symbolic and non-symbolic numerical representations draw from the same underlying semantic understanding afforded by the innate ANS (Dehaene, 2005). Put succinctly, "when we learn number symbols, we simply learn to attach their arbitrary shapes to the relevant non-symbolic quantity representations" (Dehaene, 2007, p.552).

1.2.3 Supporting Evidence for the Mapping Account

Multiple research findings have been forwarded in support of the mapping account (see Reynvoet & Sasanguie, 2016 for review). First, it has been shown that children's and adults' performance in non-symbolic number comparison (i.e., deciding which of two presented arrays of dots contains more dots) and performance in symbolic number comparison (i.e., deciding which of two Arabic numerals are larger) are highly similar in that they both exhibit a size effect and distance effect (e.g., Holloway & Ansari, 2008; Sasanguie et al., 2013). The similarity in behavior profiles of non-symbolic and symbolic number comparison has been interpreted as evidence that symbolic number understanding draws from the semantic understanding of non-symbolic numbers afforded by the ANS.

Secondly, given the hypothesized tight link between ANS ability and symbolic number ability, multiple studies have posited and have found evidence to suggest that ANS abilities are related with math achievement (e.g., DeWind & Brannon, 2012; Halberda & Feigenson, 2008). However, the nature of the relation remains ambiguous. Some studies suggest that there is a direct relation between the two variables under the premise that since both symbolic and non-symbolic numbers are represented using the same neural underpinnings, the manipulation of symbolic numbers would similarly be contingent on the neural mechanisms that allow for non-symbolic number manipulation (Dehaene, 2005; Feigenson et al., 2004; Gilmore et al., 2010). While other studies have proposed that symbolic number abilities mediate the relation between ANS ability and math achievement, and no direct relation exist between ANS ability and math achievement (Price & Fuchs, 2016).

Finally, multiple functional magnetic resonance imaging (fMRI) studies have suggested that non-symbolic and symbolic number abilities rely on the same brain area (Nieder & Dehaene, 2009). For example, using a habituation paradigm, participants are first adapted to a series of stimuli of a specific non-symbolic or symbolic quantity. When a new quantity is presented, the fMRI signal in the Intraparietal Sulcus (IPS) recovered in a distance-dependent fashion, irrespective of whether the new presented quantity is of the same notation or different notation (Piazza et al., 2007). This suggests that the IPS is agnostic to numerical notation and houses a single representation that is accessed by non-symbolic and symbolic numbers. Similarly, using a multi-voxel pattern analysis (MVPA) approach, it has been shown that patterns of activation in the parietal cortex from symbolic number stimuli can be used to predict the patterns of activation corresponding to the same non-symbolic number (Eger et al., 2009). Taken together, these fMRI experiments support the notion that symbolic and non-symbolic numbers share the same underlying representation housed in the parietal cortex.

1.2.4 Inconsistencies in Research Findings for the Mapping Account

While the mapping account had been widely accepted in the literature, recent studies testing theoretical predictions of the mapping account have found results that cannot be readily explained by the mapping account. For instance, in an experiment directly examining the link between symbolic numbers and the ANS, Lyons et al. (2012) tested participants’ ability to translate between ANS representations and symbolic number representations. Specifically, the authors compared participants' performance in a non-symbolic comparison task, a symbolic comparison task, and a mixed comparison task (deciding which of a presented Arabic numeral or array of dots is larger). Under the view of the mapping account, both symbolic and non-symbolic numbers share the same underlying representation, and translating ability (i.e., the act of translating between numerical formats when completing the mixed comparison task) would simply be an exercise of comparing two numbers on the same representation (Libertus, 2015). Therefore, performance for mixed comparison should have comparable performance to pure symbolic or pure non-symbolic comparison. However, results indicate performance for mixed comparison is significantly worse than either symbolic or non-symbolic comparison. These results call into question whether both symbolic and non-symbolic numbers draw upon the same underlying representation. Consequently, it has been argued that the additional cognitive cost of the mixed comparison task suggests that symbolic numbers do not share the same representation as non-symbolic numbers and instead, the additional cost can be seen as the time needed to connect two distinct magnitude representations to complete the task (Lyons et al., 2012).

Similarly, recent studies examining the link between ANS ability and math achievement have yielded inconsistent findings. Studies examining both children (e.g., Holloway & Ansari, 2008; Sasanguie et al., 2013) and adults (e.g., De Smedt & Gilmore, 2011; Rousselle & Noël, 2007) have failed to find a relation between non-symbolic number ability and math achievement (see De Smedt et al., 2013 for review). Critically, it has been proposed that the previously observed associations between non-symbolic comparisons and math achievement may be due to a confounding variable – inhibitory control. Specifically, the involvement of inhibitory control is a result of how continuous properties (e.g., total area, individual dot size, density, etc.) of dot arrays are experimentally controlled in non-symbolic number tasks (Clayton & Gilmore, 2015; Gilmore et al., 2013; Szucs et al., 2013). Several studies have shown the relation between ANS ability and arithmetic ability is mediated by inhibitory control (e.g., Fuhs & McNeil, 2013; Gilmore et al., 2013). Further, it was observed that the difference in performance in a non-symbolic number comparison task between children with math learning disorders and typically developing children is mainly driven by trials that most strongly depend on inhibitory control (Bugden & Ansari, 2016). Taken together, these results suggest that the link between ANS ability and math achievement may have been spurious due to a confounding variable.

Finally, recent fMRI studies have also yielded results that cannot be easily explained by the mapping account. For instance, replications of the cross-notation adaptation as reported by Piazza et al. (2007) did not find adaptation effects (Cohen Kadosh et al., 2011). Similarly, replications of the cross-notation predictivity as reported by Eger et al. (2009) were not successful (Bulthé et al., 2014). Further, using transcranial magnetic stimulation (TMS), it was observed that when the parietal regions are stimulated, non-symbolic number ability was impaired, but symbolic number ability was not impaired (Sasanguie, Göbel, & Reynvoet, 2013). In sum, recent studies have failed to replicate key findings, or have found research findings that seem to contradict the notion that non-symbolic and symbolic numbers share the same underlying neural underpinnings.

1.2.5 The Refinement Account

In light of these controversial research findings, an alternative account of numerical development has gained research attention as a potentially more accurate developmental account for the learning of symbolic numbers. The critical difference between the mapping account and the refinement account is that the refinement account posits that the central claim of the mapping account – that numeric symbols are "mapped on" to some innate ability to process numbers – is only an intermediary step that children undergo to have a full understanding of the symbolic number system.

The refinement account posits two stages of development that enables children to understand symbolic numbers. First, an initial and limited semantic understanding of symbolic numbers is acquired through a one-to-one mapping of the numbers one to four onto the object tracking system (Carey, 2001, 2004), an innate visual system that allows humans to keep track of four objects or less (Scholl, 2001; Simon, 1997). This initial understanding of small numerosities is combined with the learning of counting lists (e.g., learning to count at home or preschool) to infer characteristics of the number system (i.e., order, cardinality, and the successor function). Gradually, these characteristics are applied to numbers larger than 4, resulting in the an initial and basic understanding of the symbolic number system (Carey, 2001, 2004, 2009).

Importantly, once a basic understanding of symbolic numbers emerges, the repeated usage and practice of the system is self-reenforcing – whereby symbolic representations initially based on the object tracking system will be increasingly overwritten and reinterpreted by their relations with other numerical symbols (Mix et al., 2002; Nieder, 2009). As a result, children increasingly associate the meaning of symbolic numbers by its ordinal relation with adjacent numbers (i.e., four is after three and before five; Reynvoet & Sasanguie, 2016).

It is worth noting that under the refinement account, the ANS does not play a role in the development of symbolic number ability. It is hypothesized that the ANS is instead progressively refined as children's symbolic number abilities grow. Specifically, the processing of symbolic numbers (e.g., unrestricted precision and unconstrained upper capacity; Carey, 2004) places additional demands that requires modifications to the neural circuitries responsible for magnitude processing. Ultimately, these modifications lead to a refinement of the ANS (Verguts & Fias, 2004). Further, the learning of symbolic numbers may aid in learning characteristics of numbers that are common in both symbolic and non-symbolic formats (e.g., that numbers can be categorized into odd/even numbers), and accordingly catalyze growth in the ANS (Carey, 2004; Mix, 2008; Mussolin et al., 2014, 2016).

1.2.6 Theoretical Predictions of the Two Developmental Accounts6

As reviewed above, the two developmental accounts propose different mechanisms for the development of symbolic number ability, and consequently posit different developmental relations among the discussed precursor skills. It may be helpful to summarize the developmental predictions posited by the two developmental accounts.

Under the mapping account, ANS ability plays a central role in children's development, and three theoretical predictions are posited. First, it is proposed symbolic numerals are "mapped" on to the ANS (Dehaene, 2005). Consequently, children with stronger initial ANS abilities will be better positioned to "map on" symbolic numerals onto the ANS. Therefore, earlier ANS ability should predict later symbolic number ability. Second, given symbolic number and non-symbolic numbers draw on the same underlying representations, children who have stronger ANS ability would have an easier time retrieving the corresponding symbolic numeral attached to any particular non-symbolic representation. Therefore, children's ANS ability would be predictive of their ability to translate between formats (Geary, 2013). Finally, ANS ability has been proposed to be predictive of later math achievement, and this relation has been proposed to be mediated by symbolic number ability (Price & Fuchs, 2016), or be directly related (Dehaene, 2005; Feigenson et al., 2004; Gilmore et al., 2010).

Under the refinement account, once a basic understanding of symbolic numbers is established, the improvement of symbolic number ability is a self-reenforcing process. Three theoretical predictions are posited. First, the continual usage and practice of symbolic numbers catalyze changes in the underlying neural circuitries responsible for magnitude processing, which leads to a refinement of the ANS (Verguts & Fias, 2004). As such, earlier symbolic number ability would predict later ANS ability. Second, since continual practice of symbolic numbers stimulate the initial understanding of symbolic numerals (based on the object tracking system) to be increasingly overwritten by their relation with other symbolic numerals (Mix et al., 2002; Nieder, 2009), the efficiency in translating between numerical formats would be contingent on the rate at which the semantic number understanding matures (Lyons et al., 2018). Therefore, earlier symbolic number ability would predict later translating ability. Finally, as children who have a stronger grasp of symbolic numbers would be better positioned to learn to manipulate these numerical symbols, earlier symbolic number ability would predict later math achievement (De Smedt, Verschaffel, & Ghesquière, 2009; Durand, Hulme, Larkin, & Snowling, 2005).

1.3 Research Gaps in the Current Literature

In summary, while an impressive corpus of research has been accumulated in support of the mapping account, recent studies have either failed to replicate key supporting evidence or have found results that cannot be easily explained by the account. Consequently, an alternative account of development has been proposed as a resolution that can account for these new findings. Critically, two outstanding research questions remain in the extant literature: a) which of the two proposed developmental accounts best fits the developmental data, and b) what the reasons for the observed divergence in results are.

The main purpose of studies 1 and 2 is to contribute to answering these two research questions in the context of two under-examined issues that may have contributed to the observed divergence in results in the extant literature. By addressing these issues, studies 1 and 2 clarify the reasons why there is a divergence in research findings, and identify the developmental account that best fits developmental data.

1.3.1 Differentiating Within- and Between-Person Variability

One seldomly examined issue in the extant literature pertains to a mismatch between the theoretical predictions of the two developmental accounts and the statistical methods used to verify these theoretical predictions. More specifically, while both developmental accounts have described mechanisms for learning symbolic numbers as within-person processes (i.e., a child’s growth in symbolic number processes are contingent on some precursor skill in the same child), the statistical methods that are often used to verify these theoretical accounts provide support for between-person associations (i.e., children with stronger symbolic number abilities tends to also be children who have stronger precursor skills; Curran & Bauer, 2011; Hoffman & Stawski, 2009).

To illustrate the distinction between within-individual and between-individual variations and why this mismatch may distort research findings, it is useful to envision a case where a child’s ability in a skill (e.g., ANS or symbolic number ability) is measured across multiple timepoints. In such a scenario, the child's over-time mean – that is, the mean of a child's scores over the multiple time points – can be computed. This child-specific mean score can then be compared with the grand-mean (mean of all children's score over all time points) and yield an estimate of a child's approximate rank order ability in the population. This is referred to as between-individual variability. In contrast, a child's ability score at any specific time point can be compared with the same child's over-time mean and yield an estimate of how the child's ability changed or remain stable over time independent of their rank order in the population. This is referred to as within-individual variability.

When examined, we find that both the mapping and refinement accounts of development propose within-person mechanisms that allow for learning symbolic numbers, and therefore, are concerned with how cognitive processes change or remain the same within individuals over time. For example, under the mapping account, the learning of symbolic numbers is the act of attaching arbitrary shapes onto the corresponding non-symbolic representation (Dehaene, 2007). If the mapping account is true, we should expect a child's earlier ANS ability to be predictive of the same child's later symbolic number ability. Critically, this should be true irrespective of the child's rank order in the population.

Unfortunately, much of the statistical methods that have been commonly used to verify the implications of the mapping and refinement accounts (e.g., regressions and correlations) do not specifically address within-person variability, but rather, between-person differences. For example, if a correlation between ANS ability and symbolic number ability is found, the interpretation of this correlation is: on average, when compared with other children who have lower ANS ability, children who have higher ANS ability tends to have high levels of symbolic number ability.

The usage of between-individual test to verify within-individual accounts of development depend on the assumption that between-person relations among variables reflect the aggregated representations of within-person developmental processes (Diez-Roux, 1998). This assumption often does not hold true, and it is possible for the between-individual associations to differ both in magnitude and direction from the within-individual associations. An example from the medical literature demonstrates the need to separate within-individual and between-individual variability. It has been demonstrated that an individual has a higher probability of having a heart attack when he or she is exercising (i.e., the within-individual effect), but at the same time, people who exercise more tends to have a lower risk of heart attack (i.e., the between-individual effect; Curfman, 1993).

In the context of numerical development, as much of the supporting evidence shows between-individual associations between variables, it is currently unclear whether this observed between-individual relation is truly reflective of the hypothesized within-individual developmental processes, or some unmeasured confounding effect (e.g., from interest in mathematics, motivation, SES, etc.) mascaraing as a within-individual relation. This untested assumption behind the common methods used in the extant literature has rendered the interpretation of this body of evidence in support of one account over another a moot point. And a proper measurement of the within-person relations among the variable would be necessary to ascertain which of the two development accounts evidence support.

1.3.2 Study 1

In light of this limitation of the extant literature, the broad aim of Study 1 is to ascertain whether, when between-individual differences are controlled for, are there patterns of development that would support one developmental account over another. To this end, study 1 was a preregistered secondary data analysis with aims to determine the relations between ANS ability, symbolic number ability, translating ability, and math achievement measured over three time points. Study 1 used a random-intercept cross-lagged panel model (RI-CLPM) approach to analyze the data. The RI-CLPM has two features: 1) a random-intercept term that partials out the between-person influence, 2) a panel model that allows for the estimation of within-person developmental patterns over time. Together, the two features of the RI-CLPM allows for the estimation of within-person development processes over time and contributed to identifying the developmental account that best fits developmental data.

1.3.3 The Possibility of Subgroups in the Population

A second seldomly tested assumption of both the mapping and refinement accounts is the described mechanisms that allow for learning is equally applicable to every member in the population of children who have yet to learn symbolic numbers. In recent years, this assumption has been questioned by emergent research that have suggested distinct subgroups of children exists in the population, and different subgroups may have different developmental trajectories. For example, using a latent transition analysis technique, Chew et al. (2019) have found results suggesting that there exist subgroups of children with varying levels of association between symbolic and non-symbolic abilities – that is, some groups of children seem to exhibit a strong association between symbolic and non-symbolic abilities but other groups of children who exhibited a weaker association between the two skills (Chew et al., 2019).

The possibility of varying strengths in association between symbolic and non-symbolic number ability is challenging for both the mapping and refinement accounts. This is because both proposed developmental accounts posit a single mechanism that underlie relation between the two skills – in the case of the mapping account, earlier ANS abilities predict later symbolic number abilities, and in the case of the refinement account, earlier symbolic number abilities predict later ANS abilities. As this single mechanism should theoretically be appliable to all members of the population, the existence of subgroups with varying strengths of association would be contrary evidence suggesting a single mechanism behind the association is inadequate in describing the population. In sum, neither theoretical positions can easily explain why the strength of association between symbolic and non-symbolic number abilities differ depending on the sub-group under study.

Importantly, if subgroups with varying strengths of association exists (Chew et al., 2016, 2019), studies that failed to account for these subgroups would yield results that is the average association of all subgroups. Consequently, inconsistent research findings as outlined in the above literature review can be attributed to different research projects sampling different proportions of subgroups in their respective samples. For example, suppose that two subgroups of children exist: one group where there is a strong association between ANS ability and symbolic number ability, and another group where there is no association between the two skills. The strength of association any particular study yields will depend on the proportion of each of the subgroups that happens to be included in the study.

Considering studies pointing to the possibility of subgroups in the population, the body of literature supporting the mapping and refinement account, in its current state, may have yielded results that are misleading and may have contributed to the divergence in research findings.

1.3.4 Study 2

In light of this limitation in the extant literature, study 2 examined the possibility that there are subgroups in the population. To this end, study 2 employs a random-intercept latent transition analysis (RI-LTA) in order to ascertain whether subgroups exist in the population. The RI-LTA is composed of three features: 1) a random-intercept term that partials out the between-person influence, 2) a set of latent class variables that posits a set of underlying groups that is not observed but can be inferred from the indicator, and 3) an auto-regressive component that allows for changes in group membership over time variables (Collins & Lanza, 2009). Taken together, the RI-LTA ascertains the existence of subgroups in the population while removing the potential influences that between-person influences may exert.

1.4 Background for Study 3

The main purpose of study 3 is to contribute to the emerging issue of math anxiety and its effect on student performance. Math anxiety is an acute negative feeling that impairs sufferer’s ability to do mathematics and solve mathematical problems (Ashcraft & Faust, 1994). Math anxiety is highly prevalent across countries (Lee, 2009a) and age groups (Ma, 1999; Vukovic et al., 2013), and has been implicated to have an adverse effect on sufferer’s math achievement (Ashcraft & Krause, 2007; Barroso et al., 2020; Hembree, 1990). As a research topic, math anxiety has been examined both as a predictor of math achievement, as well as an outcome to be studied, and hopefully, alleviated.

1.4.1 Extant Literature on Math Anxiety

When viewing math anxiety as an outcome to be predicted, the extant literature seems to have identified a large repertoire of individual and environmental factors that seem to be predictive of math anxiety. These predictors include genetics (Wang et al., 2014), working memory (Ramirez et al., 2013), biases in attention (Rubinsten et al., 2015), parent's expectations and attitudes (Soni & Kumari, 2017a; Vukovic et al., 2013), teacher's anxiety (Beilock et al., 2010a; Schaeffer et al., 2020), classroom atmosphere and teacher expectation (Fast et al., 2010a; Mizala et al., 2015a).

Several lines of research have implicated math anxiety as a consistent predictor of math achievement. Indeed, several studies have found math anxiety to be detrimental to many school-related math skills. For example, problem solving (Hembree, 1990), simple arithmetic (Ashcraft & Faust, 1994) and basic number processing (Maloney et al., 2011). The most prominent account of the mechanisms behind the relation between math anxiety and math achievement is that math anxiety compromises critical cognitive resources needed for the manipulation of symbolic numbers. Specifically, it has been hypothesized that the worries and intrusive thoughts associated with math anxiety disrupts and compete for executive function resources (e.g., working memory) that is vital for mathematics problem solving (Ashcraft, 1992). Over the past decades, multiple behavioral (Ashcraft & Faust, 1994; Faust, 1996; Ramirez et al., 2013, 2016; Vukovic et al., 2013) and neuroimaging studies (Lyons & Beilock, 2011; Pletzer et al., 2015; Sarkar et al., 2014) have uncovered evidence that is consistent this view.

In sum, extant research has suggested that multiple individual and environmental factors converge to form math anxiety, and that the worries and negative thoughts that math anxiety elicits compete with mental resources needed for mathematical problem solving.

1.5 Current Research Gaps

While extant studies on math achievement have extended our understanding of the predictors of math anxiety and math anxiety's effects on math achievement, these studies are limited in three important ways that have left multiple important research questions unanswered.

1.5.1 Research Done in Isolation

First, current research into the causes of math anxiety has typically been conducted in isolation. For instance, studies examining the effects of cognitive predictors of math anxiety (e.g., Ramirez et al., 2013), have typically been conducted separately from studies examining classroom factors (Beilock et al., 2010a; Schaeffer et al., 2020). As such, it is currently unknown which of these predictors uniquely predict math anxiety and which predictors predict shared variability of math achievement. Further, studies that examine the predictors of math anxiety, and studies that examine the relation between math anxiety and math achievement have also typically been done in isolation of one another. Consequently, it is not known whether any of the identified predictors of math anxiety may serve as confounders that potentially inflate the relation between math anxiety and math achievement.

1.5.2 Multilevel Nature of Math Anxiety

Second, there currently is a mismatch between research findings, on the one hand, from the studies of the predictors of math anxiety, and on the other hand, how research is typically conducted when examining the link between math anxiety and math achievement. Specifically, multiple predictors of math anxiety have been identified to be at the nested level of the education environment (e.g., classroom atmosphere (Fast et al., 2010b) and teacher's own math anxiety (Beilock et al., 2010b; Schaeffer et al., 2020)). This suggest that there is some degree of systematic (i.e., meaningful) variations of math anxiety at the level of the education environment that may be independent of the variability at the level of the individual. Given this, when studying math anxiety as a predictor of math achievement, it is important to consider variability in individual math anxiety and variability of education environment average math anxiety as separate predictors that may exert independent and separate effects on student math achievement.

To see why this might be the case, it is useful to envision a hypothetical situation where a child with some degree of math anxiety is either put into an education environment with high average math anxiety or an education environment with low average math anxiety. In such a scenario, it is possible that education environment average math anxiety may predict student math achievement over and above what the child's own math anxiety may predict. Formally put, when the aggregation (i.e., mean) of a variable at a higher level (i.e., the education environment average math anxiety) can contribute to predicting the outcome variable (i.e., student math achievement) over and beyond what the same variable at the lower level (i.e., individual math anxiety) can predict, a contextual effect is said to have occurred.

To date, no studies have examined whether systematic variability of math anxiety at the education environment level may have a contextual effect on math achievement. It is particularly important to explore the potential of contextual effect because contextual effects could significantly differ in both size and direction when compared with the individual effect. For example, studies regarding the big-fish-little-pond effect (Marsh et al., 2008) have showed that while at the individual level, a child's relation between achievement and self-efficacy is positive (i.e., the higher a child's own grades, the higher his or her self-efficacy), at the classroom level, the relation between achievement and self-efficacy is negative (i.e., the higher a child's peers' grades, the lower his or her self-efficacy).

To the best of our ability, we have found no studies that have explored the potential of the contextual effect of math anxiety. As it is possible for contextual effects to differ both in strength and direction from the individual level effect, current understanding of how math anxiety affects math achievement remain incomplete and ambiguous.

1.5.3 A Global Perspective

Similar to other areas in psychological research, the extant literature on math anxiety suffers from an overrepresentation of research findings from western, educated, industrialized, rich and democratic (WEIRD) societies (Henrich et al., 2010). Indeed, few studies to date have systematically explored between-country differences in math anxiety (Foley et al., 2017; Lee, 2009b).

Given the added unexplored complexity of the potential of a contextual effect of math anxiety, multiple important between-country differences in how math anxiety may affect math performance remain unanswered. More specifically, there may be between-country differences in the magnitude of effect of both the individual effect (i.e., whether differences in students' own math anxiety may affect math achievement) and contextual effect (i.e., whether differences in education environment-average math anxiety may affect math achievement).

It is important to examine such between-country differences, as these differences speaks to whether research findings yielded from WEIRD countries can be generalized to other countries and cultures.

1.5.4 Study 3

The main goal of study 3 was to fill in the abovementioned research gap using a multilevel modelling framework. Study 3 utilized on data from three large-scale international studies of student achievement, 1) Trends in International Mathematics and Science Study 2015 Grade 4 (TIMSS Grade 4; Martin et al., 2016), 2) Trends in International Mathematics and Science Study 2015 Grade 8 (TIMSS Grade 8; Martin et al., 2016),, and 3) the Programme for International Student Assessment 2012 Grade 8 (PISA; OECD, 2014). Three research questions were explored in study 3.

First, few studies have explored the potential of a contextual effect of math anxiety at the education environment level, as well as few studies having systematically explored between-country differences in the individual and contextual effect of math anxiety. Study 3 will ascertain whether there is evidence for a contextual effect of math anxiety and whether there are between-country differences in the individual and contextual effects. In doing so, study 3 will address whether current research findings can be easily generalized to different cultures.

Second, much of current research on the predictors of math anxiety have been done so in isolation. The TIMSS Grade 4 include a rich set of potentially relevant predictors of math anxiety (see Chapter 4), and this allows study 3 to examine the relative strength of these predictors of math anxiety, and ascertain whether the seeming multifaceted nature of math anxiety is an artefact of how research is conducted, or is indeed a reflection of how math anxiety is influenced in reality.

Third, research looking at the predictors of math anxiety and research looking into the relation between math anxiety and math achievement have also been done in relative isolation. As such, little is known on whether any of these predictors of math anxiety may confound the relation between math anxiety and math achievement. Study 3 examined this possibility by ascertaining whether math anxiety uniquely predicts math achievement after other predictors have been accounted for in the model.

1.6 References

Agrillo, C. (2015). Numerical and arithmetic abilities in non-primate species. R. Cohen Cadosh & Ann Dowker (Eds.), The Oxford Handbook of Numerical Cognition, 214–236. https://doi.org/10.1093/oxfordhb/9780199642342.013.002

Antell, S. E., & Keating, D. P. (1983). Perception of Numerical Invariance in Neonates. Child Development, 54(3), 695–701. https://doi.org/10.2307/1130057

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. Cognition, 44(1–2), 75–106.

Ashcraft, M. H., & Faust, M. W. (1994a). Mathematics anxiety and mental arithmetic performance: An exploratory investigation. Cognition & Emotion, 8(2), 97–125.

Ashcraft, M. H., & Faust, M. W. (1994b). Mathematics anxiety and mental arithmetic performance: An exploratory investigation. Cognition & Emotion, 8(2), 97–125.

Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. Psychonomic Bulletin & Review, 14(2), 243–248.

Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2020). A meta-analysis of the relation between math anxiety and math achievement. Psychological Bulletin.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010a). Female teachers' math anxiety affects girls' math achievement. Proceedings of the National Academy of Sciences, 107(5), 1860–1863.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010b). Female teachers' math anxiety affects girls' math achievement. Proceedings of the National Academy of Sciences, 107(5), 1860–1863.

Bugden, S., & Ansari, D. (2016). Probing the nature of deficits in the 'approximate number system'in children with persistent developmental dyscalculia. Developmental Science, 19(5), 817–833.

Bulthé, J., De Smedt, B., & de Beeck, H. O. (2014). Format-dependent representations of symbolic and non-symbolic numbers in the human cortex as revealed by multi-voxel pattern analyses. NeuroImage, 87, 311–322.

Carey, S. (2001). Cognitive foundations of arithmetic: Evolution and ontogenisis. Mind & Language, 16(1), 37–55.

Carey, S. (2004). Bootstrapping & the origin of concepts. Daedalus, 133(1), 59–68. https://doi.org/10.1162/001152604772746701

Carey, S. (2009). The Origin of Concepts. Oxford University Press.

Cheries, E. W., Wynn, K., & Scholl, B. J. (2006). Interrupting infants' persisting object representations: An object‐based limit? Developmental Science, 9(5), F50–F58.

Chew, C. S., Forte, J. D., & Reeve, R. A. (2016). Cognitive factors affecting children's nonsymbolic and symbolic magnitude judgment abilities: A latent profile analysis. Journal of Experimental Child Psychology, 152, 173–191.

Chew, C. S., Forte, J. D., & Reeve, R. A. (2019). Implications of Change/Stability Patterns in Children's Non-symbolic and Symbolic Magnitude Judgment Abilities Over One Year: A Latent Transition Analysis. Frontiers in Psychology, 10, 441.

Clayton, S., & Gilmore, C. K. (2015). Inhibition in dot comparison tasks. Zdm, 47(5), 759–770.

Cohen Kadosh, R., Bahrami, B., Walsh, V., Butterworth, B., Popescu, T., & Price, C. J. (2011). Specialization in the human brain: The case of numbers. Frontiers in Human Neuroscience, 5, 62.

Collins, L. M., & Lanza, S. T. (2009). Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences (Vol. 718). John Wiley & Sons.

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. https://doi.org/10.3758/bf03196206

Curfman, G. D. (1993). Is exercise beneficial—Or hazardous—To your heart? Mass Medical Soc.

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. Annual Review of Psychology, 62, 583–619.

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

De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. Journal of Experimental Child Psychology, 103(2), 186–201. https://doi.org/10.1016/j.jecp.2009.01.004

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(2), 48–55. https://doi.org/10.1016/j.tine.2013.06.001

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

Dehaene, S. (1999). The Number Sense: How the Mind Creates Mathematics. Oxford University Press.

Dehaene, S. (2005). Evolution of human cortical circuits for reading and arithmetic: The "neuronal recycling" hypothesis. From Monkey Brain to Human Brain, 133–157.

Dehaene, S. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In Sensorimotor foundations of higher cognition (pp. 527–574).

DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: Effects of feedback and training. Frontiers in Human Neuroscience, 6. https://doi.org/10.3389/fnhum.2012.00068

Diez-Roux, A. V. (1998). Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. American Journal of Public Health, 88(2), 216–222.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., & Brooks-Gunn, J. (2007). School readiness and later achievement. Developmental Psychology, 43(6), 1428.

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

Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., & Kleinschmidt, A. (2009). Deciphering cortical number coding from human brain activity patterns. Current Biology, 19(19), 1608–1615.

Fast, L. A., Lewis, J. L., Bryant, M. J., Bocian, K. A., Cardullo, R. A., Rettig, M., & Hammond, K. A. (2010a). Does math self-efficacy mediate the effect of the perceived classroom environment on standardized math test performance? Journal of Educational Psychology, 102(3), 729.

Fast, L. A., Lewis, J. L., Bryant, M. J., Bocian, K. A., Cardullo, R. A., Rettig, M., & Hammond, K. A. (2010b). Does math self-efficacy mediate the effect of the perceived classroom environment on standardized math test performance? Journal of Educational Psychology, 102(3), 729.

Faust, M. W. (1996). Mathematics anxiety effects in simple and complex addition. Mathematical Cognition, 2(1), 25–62.

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

Foley, A. E., Herts, J. B., Borgonovi, F., Guerriero, S., Levine, S. C., & Beilock, S. L. (2017). The math anxiety-performance link: A global phenomenon. Current Directions in Psychological Science, 26(1), 52–58.

Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low - income homes: Contributions of inhibitory control. Developmental Science, 16(1), 136–148.

Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. Current Directions in Psychological Science, 22(1), 23–27.

Gilmore, C. K., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., Simms, V., & Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. PloS One, 8(6), e67374.

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.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the ‘number sense’: The approximate number system. In 3-, 4-, 5-, 6-Year-Olds and Adults. Developmental Psychology, 1457–1465. https://doi.org/10.1037/a0012682

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. Journal for Research in Mathematics Education, 33–46.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? Behavioral and Brain Sciences, 33(2–3), 61–83.

Hoffman, L., & Stawski, R. S. (2009). Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. Research in Human Development, 6(2–3), 97–120.

Holloway, I. D., & Ansari, D. (2008). Domain‐specific and domain‐general changes in children's development of number comparison. Developmental Science, 11(5), 644–649.

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first‐grade math achievement from developmental number sense trajectories. Learning Disabilities Research & Practice, 22(1), 36–46. https://doi.org/10.1111/j.1540-5826.2007.00229.x

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. Developmental Psychology, 45(3), 850. https://doi.org/10.1037/a0014939

Lee, J. (2009a). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. Learning and Individual Differences, 19(3), 355–365.

Lee, J. (2009b). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. Learning and Individual Differences, 19(3), 355–365.

Libertus, M. E. (2015). The role of intuitive approximation skills for school math abilities. Mind, Brain, and Education, 9(2), 112–120. https://doi.org/10.1111/mbe.12072

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(4), 635. https://doi.org/10.1037/a0027248

Lyons, I. M., & Beilock, S. L. (2011). Mathematics anxiety: Separating the math from the anxiety. Cerebral Cortex, 22(9), 2102–2110.

Lyons, I. M., Bugden, S., Zheng, S., De Jesus, S., & Ansari, D. (2018). Symbolic number skills predict growth in nonsymbolic number skills in kindergarteners. Developmental Psychology, 54(3), 440. https://doi.org/10.1037/dev0000445

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. Journal for Research in Mathematics Education, 520–540.

Maloney, E. A., Ansari, D., & Fugelsang, J. A. (2011). Rapid Communication: The effect of mathematics anxiety on the processing of numerical magnitude. Quarterly Journal of Experimental Psychology, 64(1), 10–16.

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K.-T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish–little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. Educational Psychology Review, 20(3), 319–350.

Martin, M. O., Mullis, I. V., & Hooper, M. (2016). Methods and procedures in TIMSS 2015. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA).

Mix, K. S. (2008). Surface similarity and label knowledge impact early numerical comparisons. British Journal of Developmental Psychology, 26(1), 13–32. https://doi.org/10.1348/026151007X189109

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). Quantitative development in infancy and early childhood. Oxford University Press.

Mizala, A., Martínez, F., & Martínez, S. (2015). Pre-service elementary school teachers' expectations about student performance: How their beliefs are affected by their mathematics anxiety and student's gender. Teaching and Teacher Education, 50, 70–78.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. Nature, 215(5109), 1519.

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. https://doi.org/10.1371/journal.pone.0091839

Mussolin, C., Nys, J., Leybaert, J., & Content, A. (2016). How approximate and exact number skills are related to each other across development: A review. Developmental Review, 39, 1–15. https://doi.org/10.1016/j.dr.2014.11.001

Nicholls, G. M., Wolfe, H., Besterfield - Sacre, M., Shuman, L. J., & Larpkiattaworn, S. (2007). A method for identifying variables for predicting STEM enrollment. Journal of Engineering Education, 96(1), 33–44.

Nieder, A. (2009). Prefrontal cortex and the evolution of symbolic reference. Current Opinion in Neurobiology, 19(1), 99–108. https://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(1), 185–208. https://doi.org/10.1146/annurev.neuro.051508.135550

Nieder, A., & Merten, K. (2007). A Labeled-Line Code for Small and Large Numerosities in the Monkey Prefrontal Cortex. Journal of Neuroscience, 27(22), 5986–5993. https://doi.org/10.1523/JNEUROSCI.1056-07.2007

OECD. (2014). PISA 2012 technical report. OECD publishing Paris.

Parsons, S., & Bynner, J. (2005). Does numeracy matter more?

Piazza, M. (2011). Neurocognitive start-up tools for symbolic number representations. Space, Time and Number in the Brain, 267–285.

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. https://doi.org/10.1016/j.neuron.2006.11.022

Pletzer, B., Kronbichler, M., Nuerk, H.-C., & Kerschbaum, H. H. (2015). Mathematics anxiety reduces default mode network deactivation in response to numerical tasks. Frontiers in Human Neuroscience, 9, 202.

Price, G. R., & Fuchs, L. S. (2016). The Mediating Relation between Symbolic and Nonsymbolic Foundations of Math Competence. PLOS ONE, 11(2), e0148981. https://doi.org/10.1371/journal.pone.0148981

Ramirez, G., Chang, H., Maloney, E. A., Levine, S. C., & Beilock, S. L. (2016). On the relationship between math anxiety and math achievement in early elementary school: The role of problem solving strategies. Journal of Experimental Child Psychology, 141, 83–100.

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. Journal of Cognition and Development, 14(2), 187–202.

Reynvoet, B., & Sasanguie, D. (2016). The symbol grounding problem revisited: A thorough evaluation of the ANS mapping account and the proposal of an alternative account based on symbol–symbol associations. Frontiers in Psychology, 7, 1581.

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. Psychological Science, 24(7), 1301–1308.

Rousselle, L., & Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. Cognition, 102(3), 361–395. https://doi.org/10.1016/j.cognition.2006.01.005

Rubinsten, O., Eidlin, H., Wohl, H., & Akibli, O. (2015). Attentional bias in math anxiety. Frontiers in Psychology, 6, 1539.

Sarkar, A., Dowker, A., & Kadosh, R. C. (2014). Cognitive enhancement or cognitive cost: Trait-specific outcomes of brain stimulation in the case of mathematics anxiety. Journal of Neuroscience, 34(50), 16605–16610.

Sasanguie, D., Göbel, S. M., & Reynvoet, B. (2013). Left parietal TMS disturbs priming between symbolic and non-symbolic number representations. Neuropsychologia, 51(8), 1528–1533.

Sasanguie, D., Göbel, S., Moll, K., Smets, K., & Reynvoet, B. (2013). Acuity of the approximate number sense, symbolic number comparison or mapping numbers onto space: What underlies mathematics achievement? Journal of Experimental Child Psychology, 114(3), 418–431.

Schaeffer, M. W., Rozek, C. S., Maloney, E. A., Berkowitz, T., Levine, S. C., & Beilock, S. L. (2020). Elementary school teachers' math anxiety and students' math learning: A large‐scale replication. Developmental Science, e13080.

Scholl, B. J. (2001). Objects and attention: The state of the art. Cognition, 80(1–2), 1–46.

Simon, T. J. (1997). Reconceptualizing the origins of number knowledge: A “non-numerical” account. Cognitive Development, 12(3), 349–372.

Skagerlund, K., Lind, T., Strömbäck, C., Tinghög, G., & Västfjäll, D. (2018). Financial literacy and the role of numeracy–How individuals’ attitude and affinity with numbers influence financial literacy. Journal of Behavioral and Experimental Economics, 74, 18–25.

Soni, A., & Kumari, S. (2017). The role of parental math anxiety and math attitude in their children’s math achievement. International Journal of Science and Mathematics Education, 15(2), 331–347.

Szucs, D., Nobes, A., Devine, A., Gabriel, F. C., & Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. Frontiers in Psychology, 4, 444.

Van Loosbroek, E., & Smitsman, A. W. (1990). Visual perception of numerosity in infancy. Developmental Psychology, 26(6), 916.

Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: A neural model. Journal of Cognitive Neuroscience, 16(9), 1493–1504. https://doi.org/10.1162/0898929042568497

Viswanathan, P., & Nieder, A. (2013). Neuronal correlates of a visual “sense of number” in primate parietal and prefrontal cortices. Proceedings of the National Academy of Sciences, 110(27), 11187–11192. https://doi.org/10.1073/pnas.1308141110

Vukovic, R. K., Kieffer, M. J., Bailey, S. P., & Harari, R. R. (2013). Mathematics anxiety in young children: Concurrent and longitudinal associations with mathematical performance. Contemporary Educational Psychology, 38(1), 1–10.

Wang, Z., Hart, S. A., Kovas, Y., Lukowski, S., Soden, B., Thompson, L. A., Plomin, R., McLoughlin, G., Bartlett, C. W., & Lyons, I. M. (2014). Who is afraid of math? Two sources of genetic variance for mathematical anxiety. Journal of Child Psychology and Psychiatry, 55(9), 1056–1064.

Williams, J., Clemens, S., Oleinikova, K., & Tarvin, K. (2003). The skills for life survey. A National Needs and Impact Survey of Literacy, Numeracy and ICT Skills. London: TSO.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. Cognition, 74(1), B1–B11.

Chapter 2

Please note that study 1 is a published manuscript entitled “Kindergarteners’ symbolic number abilities predict non-symbolic number abilities and math achievement in grade 1.” in *developmental psychology* (Lau et al., 2021).

2.1 Introduction

Mathematical competence is an important predictor of future career success, income, and well-being (Paglin & Rufolo, 1990; Parsons & Bynner, 2005b; Rivera-Batiz, 1992), and investigations of academic math achievement have suggested that variability in mathematical competency is present even at the earliest years of formal education (e.g., De Smedt, Janssen, et al., 2009; Jordan, Kaplan, Locuniak, & Ramineni, 2007; Jordan, Kaplan, Ramineni, & Locuniak, 2009). In addition to social factors such as socioeconomic status (Davis-Kean, 2005) and home learning support (Smith & Hausafus, 1998), basic numerical competencies in early childhood, such as numerical magnitude processing have been found to be predictive of student math achievement (De Smedt et al., 2013; De Smedt, Verschaffel, et al., 2009; De Smedt & Gilmore, 2011; Rousselle & Noël, 2007). Of interest to the study 1 are three different, though related, basic numerical magnitude processing competencies that underlie mathematics learning. To begin, we shall briefly describe the three basic numerical magnitude processing competencies.

First, the ability to quickly estimate the number of objects in a set is thought to be an innate cognitive system that humans possess (Dehaene, 1999; Feigenson, Dehaene, & Spelke, 2004). It has been detected in infants even a few days old (Izard et al., 2009), and has been observed in multiple animal species (Agrillo, 2015). This ability is often referred to as the non-symbolic or approximate number system (ANS) because the quantity is not represented using non-iconic symbols such as numerals. Numerical representation of the ANS is characterized by its imprecision – whereby representations are “noisy” estimates of the number of items observed following the Weber-Fechner Law. Consequently, performance at comparing the magnitude of two ANS representations are influenced by a distance effect (closer magnitudes are more difficult to compare than magnitudes farther apart; e.g., Cordes, Gelman, Gallistel, & Whalen, 2001) and a size effect (for any given distance, comparison difficulty increases with increasing average magnitude; e.g., Moyer & Landauer, 1967).

Second, many societies have systems of numerical symbols to represent quantities (e.g. number words, Arabic numerals, Chinese ideographs, etc.). These systems of number symbols are cultural inventions that rely on arbitrary (i.e., non-iconic) conventions to determine the form and rules of mathematical manipulation (Zhang & Norman, 1995). The development of children's symbolic number abilities is a protracted process that takes place over many years (Wynn, 1992). For the current study, we refer to this ability to represent and manipulate number symbols as the symbolic number system (SNS).

Third, as a consequence of humans possessing these two systems of number representation, humans can translate representations of the ANS to representations of the SNS and vice versa. The ability to translate between ANS and SNS representations incurs additional cognitive resources (Lyons et al., 2012), and this ability seems to be present even at the early stages of learning symbolic numbers (Odic et al., 2015). This ability has been referred to as the mapping between ANS and SNS (e.g., Libertus, 2015); however, we shall refer to this ability as translating ability as it carries fewer connotations regarding the relation between ANS and SNS.

2.2 Theoretical Accounts Linking Basic Numerical Skills to Future Math Achievement

Recently, there has been significant debate regarding the developmental trajectory of the ANS, SNS, and translating ability. At the heart of this discussion is whether and how these three basic magnitude processing skills are developmentally dependent on one another, and how they are related to later math achievement. Prominently, two accounts of development have been forwarded each with different predictions regarding the developmental dependency of these three skills. To empirically arbitrate between these different accounts, the current study aimed to contrast the theoretical accounts and their predictions using a longitudinal design. We shall briefly review the two contrasting theoretical accounts below.

2.2.1 The Mapping Account

Traditionally, the ANS has been postulated to form the foundation of the learning of symbolic numbers (e.g., Feigenson, Dehaene, & Spelke, 2004). This hypothesis posits that the processing of symbolic numbers co-opts, to a certain extent, the same neural circuitry for processing approximate magnitudes (Dehaene, 2005). Critically, it has been suggested that both symbolic and non-symbolic numerical representations draw from the same underlying semantic understanding afforded by the ANS. Put succinctly, 'when we learn number symbols, we simply learn to attach their arbitrary shapes to the relevant non-symbolic quantity representations' (Dehaene, 2007, p.552).

The mapping account makes three theoretical predictions regarding the development relations between ANS, SNS, translating ability, and math achievement. First, due to the semantic understanding of numerical symbols (e.g. understanding that the symbol '3' refers to all possible sets of three items) being initially affixed onto the semantic understanding afforded by the ANS, we should expect early understanding of symbolic numerals to be tightly linked with non-symbolic representations of quantities (Dehaene, 1999). Within the population of children who have yet to learn symbolic numerals, a child with stronger ANS skills would be better positioned to "map" the corresponding numerical symbols onto the ANS. Therefore, earlier ANS acuity (an individual's ability to accurately differentiate between two non-symbolic quantities) should be predictive of later SNS acuity (an individual's ability to accurately differentiate between two symbolic quantities).

Further, according to the mapping account, the ability to translate between ANS and SNS representations can be seen as the quality of mapping of symbolic numerals onto the ANS (Libertus, 2015). It has been argued that children with stronger ANS abilities would also have a stronger mapping ability between the ANS and symbolic numerals, presumably because they will have an easier time distinguishing between representations of neighboring symbolic numerals (Geary, 2013; Mussolin et al., 2012). However, the evidence is somewhat mixed with regards to whether the ANS may underlie translating ability. While some studies have found evidence to suggest there is a relation between ANS acuity and translating ability (Brankaer et al., 2014; Jang & Cho, 2018; Pinheiro-Chagas et al., 2014; Wong et al., 2016), other studies have found no such relation (Gimbert et al., 2019; Libertus et al., 2014).

Finally, it has been proposed that early ANS acuity would be predictive of later math achievement. Supporting this claim, numerous studies have shown that children and adults' earlier ANS acuity predict concurrent and later math achievement (e.g., DeWind & Brannon, 2012; Halberda & Feigenson, 2008), and training of the ANS results in improvements in symbolic calculations (e.g. Hyde, Khanum, & Spelke, 2014; Park & Brannon, 2013). The mechanism for this relation, however, remains ambiguous. One potential mechanism would be an extension of the hypothesized relation between ANS and SNS: one may predict that if early ANS acuity facilitates the acquisition of the meanings of symbolic numerals, those children with a stronger early grasp of symbolic numerals would naturally be better positioned to later acquire the skills to manipulate these numerical symbols. In other words, the predictive relation between early ANS acuity and later math achievement is mediated through the SNS (Price & Fuchs, 2016). Another proposed mechanism by which early ANS skills are related to later math achievement is through a direct route: given that multiple studies have found the ANS being capable of many basic mathematical procedures, such as relative quantity (Xu & Spelke, 2000) and basic arithmetic (McCrink & Wynn, 2009), it has been postulated that the ANS serves as a scaffold on which the understanding of these basic mathematical concepts are initially bootstrapped (Feigenson et al., 2013). Therefore, one may predict that there is instead a direct relation between ANS acuity and later math achievement.

In sum, the mapping account posited four hypotheses regarding the interactivity of early basic skills and later math achievement: (1) earlier ANS acuity predicting later SNS acuity, (2) earlier ANS acuity predicting later translating ability[1], (3) earlier ANS acuity predicting later math achievement, and (4) the potential that ANS acuity's effects on math achievement are mediated by SNS acuity.

2.2.2 The Refinement Account

An alternative view of the relation between SNS and ANS in development posits that a basic understanding (and repeated practice) of exact number symbols is self-reinforcing – that is, semantic understanding of a numerical symbol would be relatively independent of the ANS when children first learn symbolic representations of number (Le Corre & Carey, 2007), and individual symbolic representations would be increasingly understood by its relations with other numerical symbols (Mix et al., 2002; Nieder, 2009). Under this view, when one learns symbolic numbers, the additional demands of processing the unique properties of symbolic numbers (e.g., unrestricted precision and unconstrained upper capacity; Carey, 2004) would cause modifications to the neural circuitries responsible for magnitude processing that lead to the refinement of the ANS (Verguts & Fias, 2004). Further, the continual practice of symbolic numbers may promote the learning of properties intrinsic to numbers (both symbolic and non-symbolic) that may not be obvious when examined using only non-symbolic numbers (Carey, 2004; Mix, 2008; Mussolin et al., 2014, 2016). For instance, the fact that all numbers can be categorized to be either odd or even may be more easily demonstrated using symbolic numbers but would be difficult to illustrate using non-symbolic means.

Under this refinement account, one may expect the ANS and SNS to be relatively distinct at the time of initial acquisition of symbolic abilities. However, with increased practice of

[1] As reviewed above, there has been some controversy regarding whether earlier ANS acuity predicts later translating ability. The current study will proceed with the preregistered hypotheses that ANS acuity predicts later translating ability. Post-hoc analysis of mapping models that did not include this relation are included in Appendix 4. Model fit of these additional models are worse when compared with the mapping models in the main analysis.

symbolic numbers, children's ability to process non-symbolic representations will be increasingly refined by the SNS. As such, one would expect that children's earlier SNS acuity would be predictive of later ANS acuity. This claim has been theoretically bolstered by Verguts & Fias (2004), who have demonstrated in a neural network simulation study that number-selective neurons initially trained exclusively to distinguish between non-symbolic stimuli exhibited increased representational efficiency when they were next exposed to both symbolic and non-symbolic stimuli, simulating the learning of symbolic numerals. Further, several behavioral studies have suggested that children's early symbolic number abilities are predictive of ANS acuity and that the direction of influence runs from SNS to the ANS but not the other way around (Matejko & Ansari, 2016; Mussolin et al., 2014).

Further, under the refinement account, the ability to translate between ANS and SNS representations would be dependent on SNS acuity. This is because children with stronger SNS acuity would be better equipped in distinguishing between different symbolic quantities and will more easily reject a symbolic quantity that is clearly different from the one it represents. As such, a child with stronger SNS abilities will be able to arrive at the correct symbolic representation more quickly when encountering an ANS representation.

Finally, as the ability to manipulate symbolic numbers is a critical prerequisite skill for most classroom relevant mathematical tasks, children with a stronger grasp of the meaning of symbolic numbers would be better positioned to acquire skills to manipulate these symbolic numbers. As such, one might predict that early SNS ability being predictive of later math achievement. This hypothesis has been supported by multiple studies that have found that earlier SNS acuity predicted concurrent and later math achievement (e.g., De Smedt, Verschaffel, & Ghesquière, 2009; Durand, Hulme, Larkin, & Snowling, 2005).

In sum, the refinement account posits three hypotheses regarding the interactivity of early basic skills and later math achievement: (1) earlier SNS acuity predicting later ANS acuity,

(2) earlier SNS acuity predicting later translating ability[2], (3) earlier SNS acuity predicting later math achievement.

2.3 Mapping Account vs. Refinement Account

As can be seen above, the mapping account and refinement account have posited different developmental trajectories of ANS, SNS, and translating ability, and while supporting evidence for each account has been forwarded (e.g., Halberda & Feigenson, 2008; Mussolin et al., 2016), few studies have examined all the hypothesized relations between the variables simultaneously. In this context, Lyons, Bugden, Zheng, De Jesus, & Ansari (2018), attempted to disentangle the two accounts using a longitudinal design with two critical time-points (i.e., at the beginning of senior kindergarten, at approximately 5 years of age and approximately 6 months into senior kindergarten), each measuring children's early magnitude comparison abilities. This is a critical period of children's mathematical development as it is during this period in the Canadian education system when children are first transitioning from learning mostly informal mathematical knowledge to the learning of more formal mathematical knowledge (Purpura et al., 2013). Specifically, Lyons et al., (2018) addressed two main research questions: (1) To what extent are the ANS and SNS linked at the stage when formal mathematical knowledge is being introduced? (2) Does early ANS acuity predict later SNS acuity or is the relation the other way around?

It was hypothesized that if the mapping account is correct, then ANS acuity at the beginning of senior kindergarten should predict both SNS acuity and translating ability. Conversely, if the refinement account is correct, then SNS acuity at the beginning of senior kindergarten should predict both ANS acuity and translating ability. Results were most consistent with the

[2] Unlike the link between SNS acuity and ANS acuity, the link between SNS acuity and translating ability may be more theoretically tenuous. We will proceed with the preregistered hypotheses that SNS acuity predicts later translating ability. However, we have included post-hoc analysis of the Refinement Model that did not include this link are in Appendix 4. Model fit is worse when compared with the refinement model in the main analysis.

refinement account, with SNS acuity being predictive of both ANS acuity and translating ability. Importantly, even when bi-directional influences were allowed, ANS acuity did not predict SNS acuity or translating ability (Lyons et al., 2018).

2.4 Study 1

Study 1 was a preregistered secondary data analysis (see https://osf.io/xgk89/ for the preregistration documentation) with aims to replicate and extend the results by Lyons et al., (2018). While the two time-point cross-lagged design used by Lyons et al., (2018) was able to establish temporal predominance among the variables, the analysis cannot make any meaningful inferences regarding whether the model provided a proper description of the underlying interactivity among variables because a two time-point model is just-identified and cannot provide meaningful measures of model fit. Second, the cross-lagged panel design employed has been criticized as it does not distinguish between- and within-person variance (Hamaker et al., 2015). As such, it is unclear whether the regression coefficients yielded by Lyons et al., (2018) reflect a within-person process (i.e., growth in SNS abilities within a child is associated with stronger later ANS abilities within the same child) or a between-persons process (i.e., children with stronger SNS abilities compared with peers tends to have stronger ANS abilities compared with peers in the future) or a combination of both (Berry & Willoughby, 2017; Hamaker et al., 2015).

The current study addressed these issues by analyzing the relations between ANS, SNS, translating ability, and math achievement over three time-points under a random intercept cross-lagged Model (RI-CLPM) framework. Using three time-points, the current study was able to examine the stability of the relations between the variables (i.e., from T1 to T2 and from T2 to T3). Further, our analyses allowed us to directly compare the fit of competing models under the different theoretical models discussed above. Finally, the RI-CLPM specification allowed the current study to better tease apart the between- and within-person relations and to focus on the within-person changes (see **Analytic Procedures**).

2.5 Methods

2.5.1 Explanation of Existing Data

The dataset came from an ongoing collaboration between the University of Western Ontario (UWO) and the Toronto District School Board (TDSB). Study 1was a follow up on additional data that were collected. The dataset contained anonymized data of 1545 children over a period of 2 years. Variables collected include demographic information (age, birth country, school socioeconomic status (SES)), assessments from the Numeracy Screener, and grade 1 academic grades (Term 1 and Term 2). No statistical analysis or summary statistics were computed for any variables not described below.

The ongoing collaboration between the TDSB and UWO was approved by the TDSB's External Research Review Board (ERRC), and testing material was approved by the University of Western Ontario's Non-Medical Research Ethics Board (REB#102158: "The relationship between symbolic and non-symbolic numerical magnitude processing and arithmetic achievement in primary school"). Sensitive information regarding parents and students was kept confidential within the TDSB's Research and Information Services and only anonymized data were shared with researchers at UWO.

2.5.2 Participants

The children came from TDSB schools in the local area. Children were assessed over three time-points (fall 2014, spring 2015, and spring 2016; henceforth referred to as T1, T2, and T3). Mean age at T1, T2 and T3 were 62 months, 68 months, and 80 months (SD = 3.53). Of these, 628 children who complete the three critical comparison tasks (see below) for at least two of the three time-points were retained for analysis in this study. Other experimental tasks that were completed by the participants were not analyzed. Of these, five children were excluded due to misunderstandings of the task or inability to complete the tasks. The final sample was 622 children (279 females). At the time of first testing, 26 children were identified as presenting with special needs; 24 of these were learning English as a second language and 2 had unspecified special needs. As the tasks were not heavily dependent on English, these 26 children were included in the final analysis.

2.5.3 Attrition Analysis

Following Goodman and Blum's (1996) guidelines, we analyzed sample attrition across the three waves of data. A logistic regression analysis was conducted with the dependent variable being a dichotomous variable indicating whether the participant has dropped out at T2 or T3. All variables of interest at T1, time-variant covariates at T1, and time-invariant covariates were entered as predictors (Goodman & Blum, 1996). Results indicate that none of the predictors significantly predicted drop out. This suggests that at least for the variables of interest, bias due to missing data should be minimal in the model parameter estimation using full information maximum likelihood.

2.5.4 Experimental Tasks

All experimental tasks were administered using paper and pencil numeracy screener booklets (https://osf.io/xgk89/), which are based on the designs by Nosworthy, Bugden, Archibald, Evans, & Ansari (2013). At each of the three time-points, children completed the numeral comparison task, dot comparison, mixed comparison task, and an arithmetic verification task – always in that order. Each comparison task consisted of 72 total items, with 12 items per page. There is a total of 60 items for the arithmetic verification task, with 6 items per page. Children were given 2 minutes to complete as many items as quickly and as accurately as possible. They were further instructed to not skip any items. The same task items were used for each time-point.

2.5.5 Scoring

All comparison tasks and the arithmetic verification task are described below. Raw scores were computed as the total number of items completed within the two-minute time limit. To adjust for guessing, the raw scores were submitted to the equation: $A = C – I / (P-1)$, where A is the adjusted score, C is the total number of correct items, I is the total number of incorrect items, and P is the number of response options (Rowley & Traub, 1977). For the three comparison tasks and the arithmetic task, each item had two alternatives, so the adjusted score was correct minus incorrect (A = C – I). For replication purposes, this scoring method is identical to Lyons et al. (2018).

2.5.6 Numeral Comparison Task

Children's SNS was measured through performance on the numeral comparison task. In each trial of the number comparison task, children were instructed to compare two Arabic numerals arranged side-by-side and to decide which of the two numbers was numerically larger. Numbers ranged from 1-9 with the absolute distance between the two numbers (i.e., $|n_1 - n_2|$) ranging from 1 to 3. In total, 18 possible pairs of numbers were included: all 15 number pairs with distances of 1 or 2, along with 3 number pairs of distance 3 (1_4, 3_6, and 6_9). The 18 possible pairs were arranged pseudo-randomly over four blocks, giving a grand total of 72 trials. The first block of 18 pairs was arranged so that 9 pairs had the larger numeral was on the left side, and the other 9 pairs had the larger numeral on the right side. The 9 trials were chosen such that the larger side was not related to numerical size, distance, or ratio. The next 18 were arranged in the opposite manner. The last 36 trials were determined in the same manner as the first two blocks. Trial order within each block of 18 such that, for an n^{th} item in the sequence, the average numerical ratio, size, and distance were the same across the three comparison tasks (numeral, dot, and mixed). This arrangement was done so that the size and distance ratios encountered on each task would not have significantly differed across tasks (all *p*s >.20), and thus not confound with the stimulus format.

2.5.7 Dot Comparison Task

Children's ANS was measured through performance on the dot comparison task. In each trial of the dot comparison task, children were instructed to compare which of two arrays of dots arranged side-by-side and decide which of the two arrays had more dots. The children were also instructed to use their best guess and not to count the dots. Numerosities and trial order were the same as the numeral comparison task. Additionally, two versions of the task were given. In one version, dot area positively correlated with numerosity, and overall contour length was negatively correlated with numerosity, the opposite was true in the other version. As such, relying on any single parameter would have led to chance performance. Version order was also pseudo-randomized so that it was not informative of the correct answer within a given segment of trials.

2.5.8 Mixed Comparison Task

Children's translating ability was measured through performance on the mixed comparison task. In each trial of the mixed comparison task, children were instructed to compare an array of dots against an Arabic numeral arranged side-by-side and decide which of the two were numerically larger. The children were also instructed to use their best guess and not to count the dots. Numerosities and trial order were the same as the numeral comparison task. Further, the side which contained the numeral and which the dots were pseudo-randomized so that it was not informative of the correct answer.

2.5.9 Arithmetic Verification Task

Table 1. *Description of the 6 tiers of Difficulty in the Arithmetic Task*

Difficulty Tier	Addend 1	Addend 2	Outcome	Number of Problems
Tier 1	1-4	1-4	2-4	6
Tier 2	1-4	1-4	5-9	10
Tier 3	5-9	1	>6	10
Tier 4	5-9	2-4	2-10	12
Tier 5	5-9	2-4	>7	18
Tier 6	5-9	5-9	>9	4

Two math achievement measures were used in this pre-registered secondary data analysis: arithmetic ability and grade 1 grades (see below). We shall first cover arithmetic ability in this section. The operationalization of children's math achievement through arithmetic ability is in line with other studies that substitute arithmetic achievement as math achievement (e.g., Lonnemann, Linkersdörfer, Hasselhorn, & Lindberg, 2011; Mundy & Gilmore, 2009). Children's arithmetic ability was measured through an arithmetic verification task. In each arithmetic problem, children were shown an arithmetic equation (e.g., 1 + 3 = 4) and were instructed to verify whether the arithmetic equation is correct by underlining a checkmark for correct answers or an "x" for incorrect answers. These problems were organized into 6 tiers of questions of increasing difficulty (see **Table 1**). The increase in difficulty is in line with other standardized arithmetic assessments (e.g., De Vos, 1992; Woodcock, Johnson, & Mather, 1990).

Problems within each tier were arranged in a pseudo-random order. For problems in all tiers except for tier 5, problems were counter-balanced to have equal numbers of correct and incorrect arithmetic equations. Further, for incorrect trials, the foil (i.e., the incorrect sum) was larger (+1 for 1 digit outcome and +2 for 2 digit outcomes) than the true outcome for half of the incorrect trials, and smaller (-1 for 1 digit outcome and -2 for 2 digit outcomes) than the true outcome for the rest of the incorrect trials. Because there are 18 problems in total in tier 5, there is one more smaller than true outcome foil than a larger than true outcome foil in that tier. Finally, for problems in tier 1 to 5, addends are counterbalanced so that half the trials presented the smaller addend first, while the other half of the trials presented the larger addend first. For tier 6, there are three smaller addend first problems and one larger addend first problem. All trials are addition problems drawn from the problem pool of all possible combinations of addends in the range of digits 1 to 9. In total, there were 60 arithmetic problems.

Tier 1 problems consisted of addends that are within the subitizing range (1-4) whose outcomes were also within the subitizing range (1-4). Tier 2 problems contained addends within the subitizing range whose outcomes are above the subitizing range (5-9). Tier 3 was comprised of problems with one addend being a digit above the subitizing range (5-9) and the other addend being 1. Tier 4 included problems with one addend being a digit from above the subitizing range and the other addend being in the range of 2-4 whose outcomes are 10 or below. Tier 5 contained addends the same as from tier 4, except it includes outcomes above 10 as well. Finally, tier 6 included 4 randomly chosen problems from the problem pool that were not members of tiers 1-5.

2.5.10 Grade 1 Math Achievement

The second math achievement measure used in this study was children's mathematics report card grades for grade 1 school term 2. School term 2 report card grades were delivered in June 2016 and represent academic achievement from February 2016 to June 2016 (the school term immediately after T3 data collection). Grade 1 mathematics report card grades are calculated according to the guidelines as prescribed by the Ontario Ministry of Education and Training. Grades are separated into evaluations of five different strands: Number Sense and Numeration, Measurement, Geometry and Spatial Sense, Patterning and Algebra, and Data Management and Probability. Achievement in each of these strands are reported from a scale of 1 – 4: level 1 represents achievement that falls much below the provincial standard, level 2 represents achievement that approaches the provincial standard, level 3 represents the achievement of provincial standard and level 4 represents achievement that surpasses the provincial standard (Ontario Ministry of Education and Training, 2005). Students' academic achievement is operationalized as the mean score of the five strands for each student if at least three of the five strands did not contain missing data.

2.5.11 Covariates

Time-invariant covariates include age in months, gender (0: female, 1: male), whether the child was born in Canada (0: no, 1:yes), school SES and the Early Development Instrument (EDI; Janus et al., 2007).

The SES for each child was not available and SES was estimated for each school instead. Schools were classified as 1 = Low-SES (25%), 2 = Medium-Low-SES (32%), 3 = Medium-High-SES (32%), and 4 = High-SES (10%). The classification has been provided by the TDSB and is a score computed from the following variables: median household income, percentage of families below the low-Income measure (a metric of household income used by the Canadian government; Giles, 2004), percentage of families on social assistance, percentage of guardians without a high school diploma, percentage of guardians with at least one university degree, and percentage of lone-parent families.

The EDI contained sub-scores that assess a child's school readiness in five general domains, physical health and well-being, social competence, emotional maturity, language and cognitive development, and communication skills and general knowledge. Scores in each subscale range from 0-10, with high scores representing greater readiness and lower scores representing lower readiness.

Additionally, percentage of days absent during the school year (2014: M = 0.11, SD = 0.09; 2015: M = 0.09, SD = 0.07 and 2016: M = 0.07, SD = 0.06) and testing interval (in days) were included as covariates. Following Lyons et al., (2018), rather than excluding a child if they failed to perform at above chance level (adjusted score < 0), a dummy variable (coded as 1 if a child's adjusted score was less or equal to 0, or 0 if a child's adjusted score was above zero) was created for each task in for T1 and T2. The inclusion of dummy variables in T3 as covariates caused instabilities in model fit due to only a few participants (n < 20) yielded negative scores at T3 (see **Appendix 1** for more details). Therefore, for all tasks in T3, no dummy scores were included.

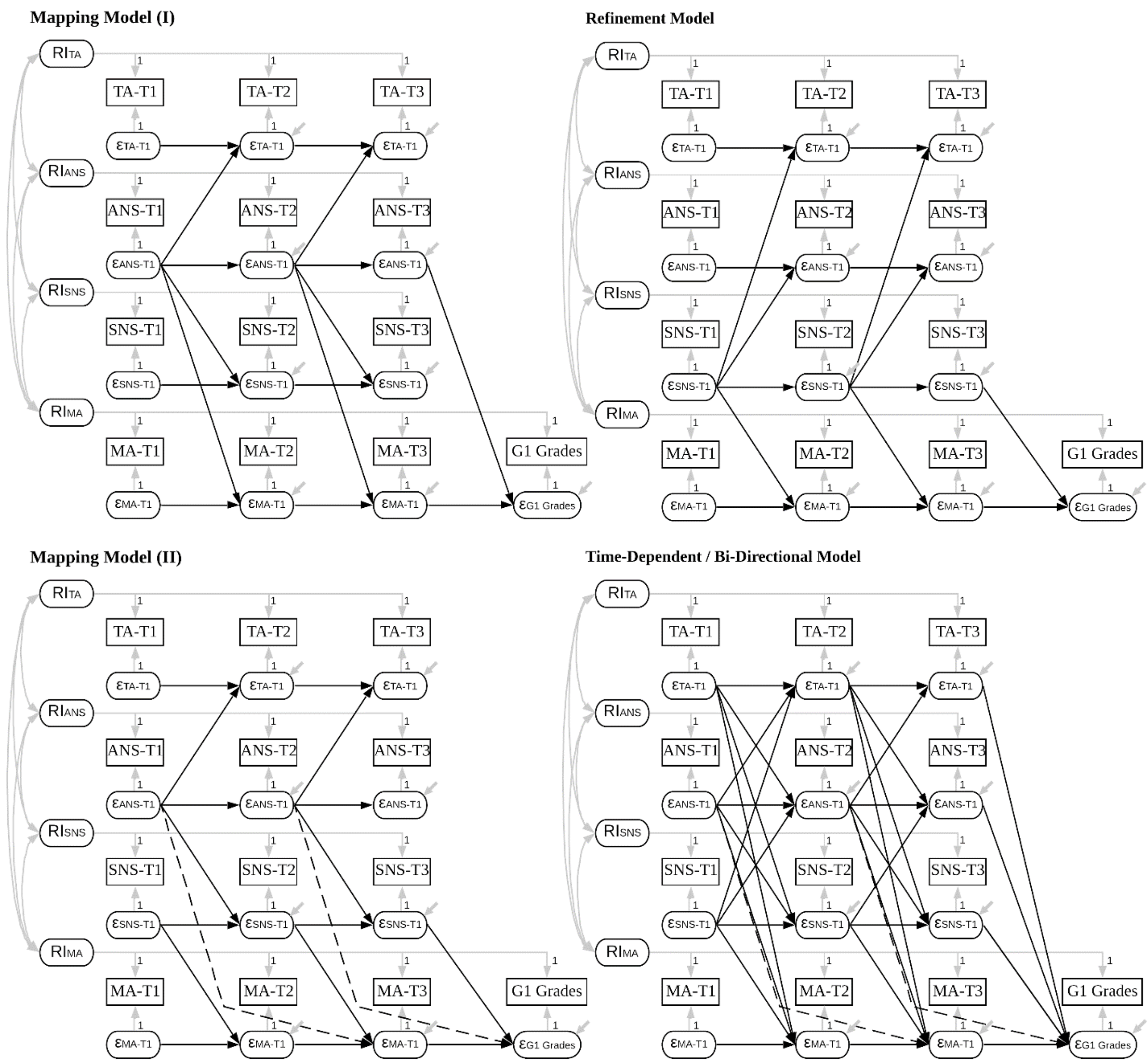
All time-varying and invariant covariates are included as predictors of the structured residuals of the main variables in their respective time-points. While the random intercepts would theoretically remove the influences of time-invariant covariates, the inclusion of time-invariant covariates at each time-point will allow for the effects of these covariates to change over time.

2.5.12 Data Treatment

Outlier detection and removal followed the guidelines laid out by Aguinis, Gottfredson, & Joo (2013) in detecting model fit outliers. Potential outliers are identified by both univariate and multivariate approaches. First, observations above or below 2.24 standard deviation units for all variables of interest were flagged as potential outliers (M. A. Martin & Roberts, 2010). Next, Mahalanobis distance was used to identify potential multivariate outliers as recommended by Becker & Gather (1999). Once potential univariate and multivariate outliers were identified, they were iteratively removed case-wise from the dataset and checked if the removal of a case changes the statistical interpretation of overall model fit. Examination of the changes in model fit has revealed that no outliers affect model fit, therefore, all participants were retained for analysis.

For all analyses, missing data were treated using full information maximum likelihood (Enders, 2001; Muthén & Muthén, 2010). Due to vastly different measurement scales between continuous variables, the variance of continuous variables was restricted to be between one and ten by scaling the variables by a constant. This practice does not affect individual differences or associations between different variables and is generally a common procedure in SEM (Little, 2013).

Figure 1. Four RI-CLPM Tested in the Main Analysis



Note. TA refers to translating ability and MA refers to math achievement. T1-3 refers to time-points of data collection. Grey lines are specifications for the random intercepts and structured residuals. Black lines are the autoregressive and cross-lagged coefficients of interest. The dotted line in Mapping (II) model and Time-Dependent / Bi-Directional model represents a direct path of mediation. Within each time-point, ANS, SNS, translating ability and math achievement residuals are allowed to co-vary. Note that no covariates (described above), residuals, or covariations are illustrated for the sake of clarity.

2.5.13 Analytic Procedures

The main analysis was performed to contrast the two accounts of numerical development as outlined above. We shall first outline the hypotheses that would be tested and contrasted. The first four hypotheses correspond to the mapping view that symbolic number understanding is predicated on approximate magnitudes (see **Figure 1** for a graphical representation of the hypotheses):

Ia. Earlier ANS predicting later SNS

Ib. Earlier ANS predicting later translating ability

Ic. ANS predicting later math achievement

Id. There may be a mediating relation between ANS and math achievement through SNS

The next three hypotheses correspond to the refinement view that symbolic number understanding sharpens and refines innate approximate magnitudes (see **Figure 1** for a graphical representation of the hypotheses):

IIa. Earlier SNS predicting later ANS

IIb. Earlier SNS predicting later translating ability

IIc. Earlier SNS predicting later math achievement

These hypotheses were explored by comparing four random-intercept cross-lagged models (RI-CLPM; **Figure 1**). Compared with the conventional cross-lagged path model, the RI-CLPM distinguishes variance at the within-level and variance at the between-level by way of estimating a random intercept for each construct in examination. This holds the advantage that time-invariant trait-like individual differences (i.e., between-persons variance) would be controlled for (Hamaker et al., 2015). Therefore, paths can be considered the degree of within-person carry-over effects across time-points while controlling for time-invariant individual differences. Specifically, path coefficients can be understood as how within-individual deviations (e.g., from growth or training) independent from one's rank order relative to other peers is predictive of a subsequent change in the same or different construct

(Hamaker et al., 2015). For example, a positive autoregressive coefficient from SNS -T1 to SNS -T2 would indicate that when an individual's SNS -T1 score is higher than expected based on the individual's symbolic number ability relative to peers (i.e., removing the influence of individual differences) the same individual's SNS score would also be higher for T2. Similarly, a positive cross-lagged coefficient from ANS-T1 to SNS -T2 would indicate that when an individual's ANS-T1 score is higher than expected based on the individual's ANS abilities relative to peers, the same individual's SNS score would be higher for T2.

Mapping Model (I) reflected the prediction that there is a direct relation between early ANS skills and later math achievement, specifically testing hypotheses Ia, Ib, and Ic. Mapping Model (II) reflected the prediction of a potential mediating relation between ANS and math achievement through SNS, specifically testing hypothesis Id in addition to hypotheses Ia - Ic. The refinement model reflected the prediction that early SNS being predictive of later math achievement, specifically testing hypotheses IIa, IIb, and IIc. Finally, a time-dependent/bi-directional model was included. This model allowed for the interrelations between ANS, SNS, and translating ability to bi-directionally affect each other, as well as allowed for time-dependent effects.

Model fit for all models was evaluated via chi-square test of fit (χ^2), Comparative Fit Index (CFI), Tucker-Lewis index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standard Root Mean Square Residual (SRMR). Preregistered thresholds for excellent fit were CFI > 0.95, TLI > 0.95, RMSEA < 0.06, and RMSR < 0.08 (Hu & Bentler, 1999). Modification indices were examined to find previously unaccounted links between variables. Once final models were computed, comparison between the models was done via comparing the Akaike Information Criterion (AIC), Bayes Information Criterion (BIC), and through an extension of Vuong's likelihood ratio test (Merkle et al., 2014; Vuong, 1989).

2.7 Results

The above-described analyses had been preregistered prior to the actual analysis being performed (https://osf.io/xgk89/), however, there were some departures in the actual analytical procedure performed and the analytical procedure described in the preregistration document. Please refer to Appendix 1 for documentation of the differences and justifications for these departures. Further, two variations in the analyses (differences in covariates used and scoring method) were preregistered in the interest of replicating the results of Lyons et al., (2018). Model fit between these variations was generally very similar, however, had some minor departures in statistical interpretations in path coefficients (see Appendix 2 a description of the variations as well as fit indices and path coefficients for all variations). The following analyses were performed using Lavaan (Rosseel, 2012) and Mplus 8.4 (L. K. Muthén & Muthén, 2010).

Table 2. Correlation and Covariance Matrix for the Variables of Interest

Variables	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. Number Comparison T1	**9.403**	0.706	0.449	0.707	0.589	0.453	0.738	0.621	0.425	0.530	0.620	0.550	0.518
2. Number Comparison T2	6.637	**9.636**	0.551	0.536	0.721	0.511	0.620	0.778	0.518	0.485	0.655	0.570	0.508
3. Number Comparison T3	3.971	5.005	**8.489**	0.371	0.496	0.755	0.457	0.512	0.743	0.284	0.406	0.655	0.513
4. Dot Comparison T1	4.029	3.022	2.040	**3.452**	0.533	0.418	0.694	0.509	0.370	0.460	0.495	0.415	0.479
5. Dot Comparison T2	3.798	4.701	2.982	2.06	**4.406**	0.467	0.523	0.774	0.494	0.436	0.592	0.441	0.413
6. Dot Comparison T3	3.055	3.492	4.835	1.749	2.117	**4.829**	0.486	0.478	0.750	0.348	0.389	0.639	0.505
7. Mixed Comparison T1	4.654	3.877	2.696	2.648	2.267	2.203	**4.224**	0.593	0.454	0.531	0.538	0.502	0.498
8. Mixed Comparison T2	4.405	5.545	3.352	2.163	3.734	2.351	2.807	**5.246**	0.521	0.492	0.657	0.532	0.462
9. Mixed Comparison T3	2.988	3.715	4.966	1.618	2.350	3.781	2.126	2.696	**5.265**	0.359	0.450	0.695	0.487
10. Arithmetic Verification T1	2.052	1.864	0.992	1.074	1.153	0.934	1.379	1.433	0.988	**1.593**	0.662	0.426	0.393
11. Arithmetic Verification T2	3.380	3.525	2.001	1.615	2.158	1.446	1.960	2.615	1.758	1.476	**3.020**	0.530	0.506
12. Arithmetic Verification T3	3.482	3.680	3.951	1.634	1.885	2.905	2.125	2.473	3.301	1.077	1.861	**4.286**	0.639
13. Grade 1 Term 2 Grades	2.072	2.024	1.910	1.169	1.088	1.418	1.320	1.348	1.429	0.634	1.128	1.689	**1.686**

*Note***.** The upper triangle refers to correlations between variables, and the bottom triangle refers to the covariances. Also, note the covariance is calculated based on the rescaled variables (see Appendix 1 for details). Due to limited space, covariates are not included in the above table, please go to https://osf.io/xgk89/ for the complete correlation/covariance matrix.

2.7.1 Descriptive Statistics

Table 3 presents the means, standard deviations, ranges, skewness, and kurtosis of the variables of interest and covariates. Examination of skewness and kurtosis revealed most of the variables modestly deviate from normality (Skewness < 2, Kurtosis < 7; Finch, West, & MacKinnon, 1997). Further examination of the variance-covariance matrix (**Table 2**) revealed that the three comparison tasks and the arithmetic verification task tend to have high correlations (maximum 0.774), which may indicate the presence of multicollinearity. As such, variance inflation factor (VIF) was calculated for all variables. Results indicate no evidence for multicollinearity (all VIFs < 5; Sheather, 2009).

The adjusted score (see **Scoring**) for all variables of interests tended to have positive skewness – indicating a concentration of scores that are below the mean. The only exception is T3 number comparison with negative skewness – indicating scores tended to be concentrated above the mean. Visual inspection of the distribution of scores showed the concentration of adjusted scores to be generally above 0, suggesting little evidence of a floor effect. Therefore, for dot comparison, mixed comparison, and arithmetic verification, the consistent positive skewness over the three time-periods seemed to reflect the nature of the distribution of these variables in the population. However, for number comparison, the shift from positive to negative skewness over the three time-periods may reflect a ceiling effect with some students completing the whole section before the expiration of the two-minute time limit. The ML estimator is robust to modest departures in normality, therefore, analyses were performed using the ML estimator as preregistered (Little, 2013).

Table 3. Descriptive Statistics of Main Variables in Study 1

Variable	**Mean**	**SD**	**Range**	**Skew**	**Kurtosis**
Number Comparison T1	19.10	15.33	-12 – 72	0.68	0.30
Number Comparison T2	30.85	15.52	-27 – 72	0.01	0.04
Number Comparison T3	45.98	14.57	-2 – 72	-0.44	0.36
Dot Comparison T1	15.06	9.29	-14 – 62	0.40	1.81
Dot Comparison T2	22.24	10.50	-12 – 69	0.72	2.36
Dot Comparison T3	30.51	10.99	-6 – 70	0.33	0.57
Mixed Comparison T1	12.96	10.28	-12 – 56	0.36	0.25
Mixed Comparison T2	21.05	11.45	-8 – 72	0.69	1.78
Mixed Comparison T3	31.15	11.47	-4 – 68	0.21	0.85
Arithmetic Verification T1	3.79	6.31	-19 – 28	1.04	2.10
Arithmetic Verification T2	9.11	8.69	-11 – 50	0.88	1.40
Arithmetic Verification T3	18.69	10.35	-13 – 53	0.10	-0.17
Grade 1 Term 2 Grades (Mean)	3.02	0.65	1 – 4	-0.44	-0.29
EDI Physical Well Being	8.72	1.35	3.85 – 10	-1.08	0.66
EDI Social Competence	8.23	1.94	1.54 – 10	-1.13	0.36
EDI Emotional Maturity	7.94	1.66	1.67 – 10	-0.93	0.56
EDI Language Cognitive	8.89	1.57	2.31 – 10	-1.89	3.18
EDI Communication General Knowledge	7.68	2.61	0 – 10	-0.93	-0.15
Assessment Interval (days) T1 – T2	192.04	14.10	141 – 217	-2.17	5.30
Assessment Interval (days) T2 – T3	382.06	10.10	351 – 426	0.51	3.82
Age in Months at T1	62.02	3.53	56.05 – 78.02	0.14	-0.63
Absent Percentage T1	0.11	0.09	0 – 0.71	1.97	5.76
Absent Percentage T2	0.09	0.07	0 – 0.5	1.82	5.06
Absent Percentage T3	0.07	0.06	0 – 0.59	2.28	10
Gender	0.55	0.50	0(45%), 1(55%)	-	-
Birth Place	0.88	0.33	0(12%), 1(88%)	-	-
SES	2.28	0.95	1(25%), (32%), 3(32%), 4(10%)	-	-
Number Comparison Dummy T1	0.09	0.28	0 – 1	-	-
Number Comparison Dummy T2	0.03	0.16	0 – 1	-	-

Dot Comparison Dummy T1	0.05	0.22	0 – 1	-	-
Dot Comparison Dummy T2	0.02	0.13	0 – 1	-	-
Mixed Comparison Dummy T1	0.11	0.31	0 – 1	-	-
Mixed Comparison Dummy T2	0.03	0.17	0 – 1	-	-
Arithmetic Verification Dummy T1	0.31	0.46	0 – 1	-	-
Arithmetic Verification Dummy T2	0.14	0.35	0 – 1	-	-

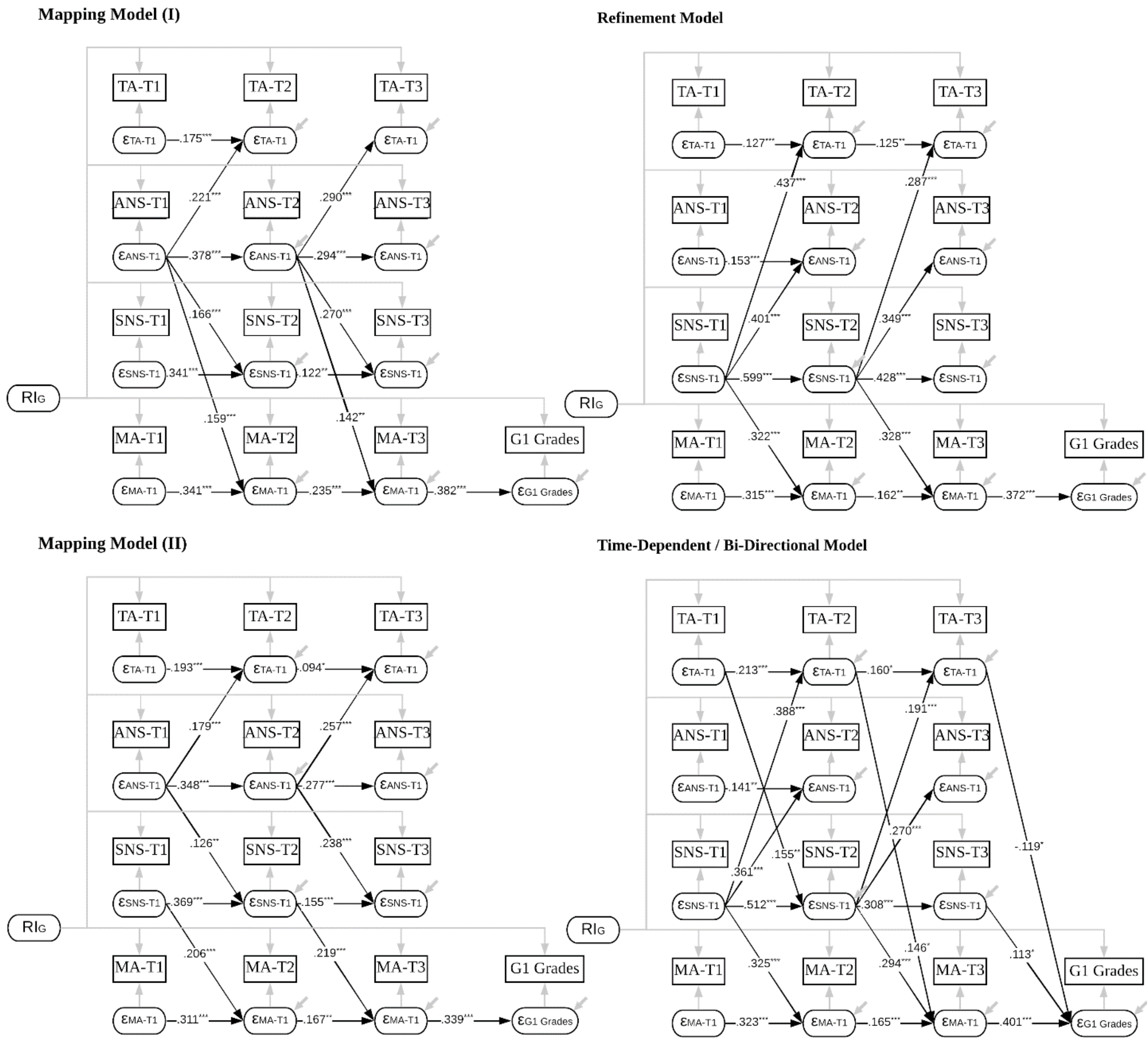
Note. Descriptive statistics of the variables and covariates (before rescaling). Skewness and kurtosis were not calculated for categorical variables, instead proportions are given in range. T1 – T3 refers to time 1 to time 3. Scores for all comparison and verification tasks are adjusted scores (see **Scoring**). Kurtosis is generally high for absent percentage, but these are likely a reflection of the nature of the population and thus not transformed.

2.7.2 Preregistered Main Analysis - The Predictive Relations between ANS, Symbolic Number Ability, Translating Ability, and Math Achievement

Four RI-CLPM models were computed and compared using the maximum likelihood estimator. Interestingly, when separate random intercepts are specified for ANS, SNS, translating ability, and math achievement, convergence problems were encountered due to the correlations between the random intercepts approaching unity. This suggested that the time-invariant trait-like individual differences that underlie ANS, SNS, translating ability, and math achievement – stable environmental and individual factors – are substantially overlapped. A potential solution is to specify a general random intercept for all four variables. Unfortunately, it is not possible to test for the appropriateness of a general random intercept for all four variables due to non-convergence. However, a model with three main variables (ANS, SNS, and translating ability) each with separate random intercepts did converge. Results indicate that the intercorrelations of the random intercepts are very high (0.968-0.980). Importantly, when this model is tested against an alternative model with a general random intercept, results indicate that the specification of a general random intercept did not reduce model fit, $\chi^2_{(5)} = 1.238$, $p = 0.941$. As such, for all subsequent analyses, instead of modeling separate random intercepts for each of the constructs, one random intercept was employed for all the constructs (see **Figure 2**).

Results indicate that the refinement and time-dependent/bi-directional models have acceptable fit, and Mapping Model (I) and Mapping Model (II) near-acceptable fit (Hu & Bentler, 1999; see **Table 4** for fit indices of each model).

Figure 2. Standardized Coefficients for the Four Main Models



Note. TA refers to translating ability and MA refers to math achievement. T1-3 refers to time-points of data collection. Grey lines are specifications for the random intercepts and structured residuals. Black lines are the autoregressive and cross-lagged coefficients of interest. Statistically insignificant path coefficients are not shown for clarity. Grey lines are specifications for the random intercepts and structured residuals and are not numerated for clarity.

Comparison between models was first conducted using AIC and BIC. Lowest AIC value was observed for the Time-Dependent/Bi-Directional Model, and the lowest BIC value was observed for the Refinement Model (**Table 4**). The models were further compared using the chi-square difference test for nested models, and using an extension of Vuong's LRT (Merkle et al., 2014; Vuong, 1989) for non-nested models. As such, Mapping Model (I), Mapping Model (II), and

Refinement Model were all tested against the Time-Dependent/Bi-Directional Model using the chi-square difference test, and the three models tested against each other using the Vuong's LRT.

Table 4. Fit Indices of the Four Main Models

	Mapping Model (I)	Mapping Model (II)	Refinement	Time-Dependent/ Bi-Directional
df	184	182	184	168
χ^2	527.939, $p <$ 0.001	489.502, $p <$ 0.001	390.219, $p <$ 0.001	348.580, $p <$ 0.001
AIC	50320.010	50285.573	50182.290	50172.651
ΔAIC	147.359	112.922	9.639	-
BIC	52452.254	52426.683	52314.534	52375.822
ΔBIC	137.72	112.149	-	61.288
CFI	0.949	0.954	0.969	0.973
TLI	0.899	0.909	0.939	0.942
RMSEA	0.055	0.052	0.042	0.042
$RMSEA_{Lower}$	0.049	0.047	0.037	0.035
$RMSEA_{Upper}$	0.060	0.058	0.048	0.048
SRMR	0.049	0.048	0.039	0.036

Note: $RMSEA_{Lower}$ and $RMSEA_{Upper}$ reflect the 90% CI. ΔAIC and ΔBIC refer to the absolute difference of model AIC and BIC with the minimum AIC and BIC of all models considered. Note that all reported models have included covariates that are not illustrated in figure 1. Please refer to covariates for more information.

Results indicate that models Mapping Model (I) and Mapping Model (II) are distinguishable ($\omega^2 = 0.109$, $p < 0.001$), and that Mapping Model (II) was statistically better fitting (z = -2.181, $p = 0.01$). Comparing Mapping Model (I) with the Refinement Model found that the two models are both distinguishable ($\omega^2 = 0.495$, $p < 0.001$) and that the Refinement Model was statistically better fitting (z = .876, $p < 0.001$). Comparing Mapping Model (II) with the Refinement Model found that the two models are both distinguishable ($\omega^2 = 0.434$, $p < 0.001$) and that the Refinement Model was statistically better fitting (z = 3.044, $p = 0.001$).

Results from three separate chi-square difference tests indicate that the Time-Dependent/Bi-Directional Model has superior model fit when compared with Mapping Model (I) ($\chi^2(16) = 179.36$, p < 0.001), Mapping Model (II) ($\chi^2(14) = 140.92$, p < 0.001) and the Refinement Model ($\chi^2(16) = 41.63$, p < 0.001). In sum, an examination of AIC and the chi-square difference test revealed the Time-Dependent/Bi-Directional Model to be most superior, and examination of BIC found the Refinement Model to be most superior (see **Figure 2**).

Comparison of the four models revealed that the Time-Dependent/Bi-Directional Model and Refinement Model have the most superior model fit. Examination of the significant paths of the Refinement model revealed earlier SNS is consistently predictive of later ANS acuity, translating ability, and math achievement across T1 to T3.

As the Time-Dependent/Bi-Directional Model allowed for both bi-directional and time-dependent effects, the statistical significance of the path coefficients in the model was examined. Notably, other than three additional transient paths related to translating ability, the Time-Dependent/Bi-Directional Model is in perfect agreement with the Refinement Model. Therefore, these results are highly supportive of the notion that earlier SNS is consistently predictive of later ANS acuity, translating ability, and math achievement, supporting hypotheses IIa – IIc.

The departures between the two models are transient effects pertaining to the predictive ability of translating ability. The transient effects of translating ability may be due to the operationalization of translating ability of the study 1. Preschool children's ability to translate from symbolic number representations to ANS representations seems to be stronger than their ability to translate from ANS representations to symbolic number representations (Odic et al., 2015). Since the mixed comparison task does not impose restrictions on the direction of translation, the longitudinal measurement of mixed comparison over three time-points may be capturing different proportions of variance contributed by the two directions of translation as children develop over time. This suggests that translating ability may play a role in children's ANS and SNS development and that the basic skills that predict translating ability and the basic skills that translating ability predicts may differ depending on the direction of translation.

2.7.3 Choices in Model Selection

In the main analysis, we found the Time-Dependent/Bi-Directional Model and Refinement Model to be the best fitting models. We found that AIC and the chi-square difference test favored the Time-Dependent/Bi-Directional model, and BIC favored the Refinement model. The absolute difference in AIC or BIC values can be used to gauge the relative probability that one model has a better fit than another and ΔAIC and ΔBIC > 10 is considered substantial posterior odds in favor of the model with the lower AIC or BIC value (Burnham & Anderson, 2003; Raftery, 1995).

Therefore, in our results, AIC somewhat favors the Time-Dependent/Bi-Directional Model, and BIC strongly favors the Refinement Model.

As AIC and BIC differ in preference, we cannot make a definite statement regarding which of the two best-fitting models is superior. However, we could offer our opinion: generally, AIC has the tendency to favor models that are overfitted to random variations of the sample, while BIC has a tendency to favor the more parsimonious model and may be biased in overlooking weaker relations (Kuha, 2004). Therefore, which is the "best" fitting model would depend on the purpose these models serve. For the purposes of generalizing to the population, the Refinement Model would probably be best, and for the purposes of studying the relations between the variables in more detail, the Time-Dependent/Bi-Directional Model would be a good starting point.

2.7.4 Post-Hoc Analysis - Translating Ability as a Predictor under the Mapping Account

Under the mediated mapping account, growth in children's SNS was proposed to be a mediating variable in the relation between ANS and math achievement. The rationale behind this mediating relation is that those children with a stronger ANS acuity will be better positioned to "map on" symbols onto the ANS, and subsequently, those children with a stronger grasp of symbolic numerals would be better positioned to acquire school-taught mathematical skills. While some researchers have explored this relation using SNS as the mediator (as we have done in the current study; e.g., Price & Fuchs, 2016), other researchers have explored this relation using translating ability as the mediator (e.g., Libertus et al., 2016), or both SNS acuity and translating ability as dual mediators (e.g., Brankaer et al., 2014). Finally, there is the possibility that translating ability and SNS work as sequential mediators (i.e., ANS acuity > Translating Ability > SNS acuity > Math Achievement), as it could be reasoned from the mapping perspective that students who have stronger mapping of symbols onto the ANS may have an easier time learning symbol-to-symbol relations.

These possible additional mediating relations are represented in Mapping Model (III - V) and were compared to the two best-fitting models (see **Figure 3**). Mapping Model (III) corresponds to the hypothesis that translating ability serves as the sole mediator. Mapping Model (IV) corresponds to the hypothesis that both translating ability and SNS serves as dual mediators. Mapping Model (V) corresponds to the hypothesis that translating ability and SNS serves as sequential mediators.

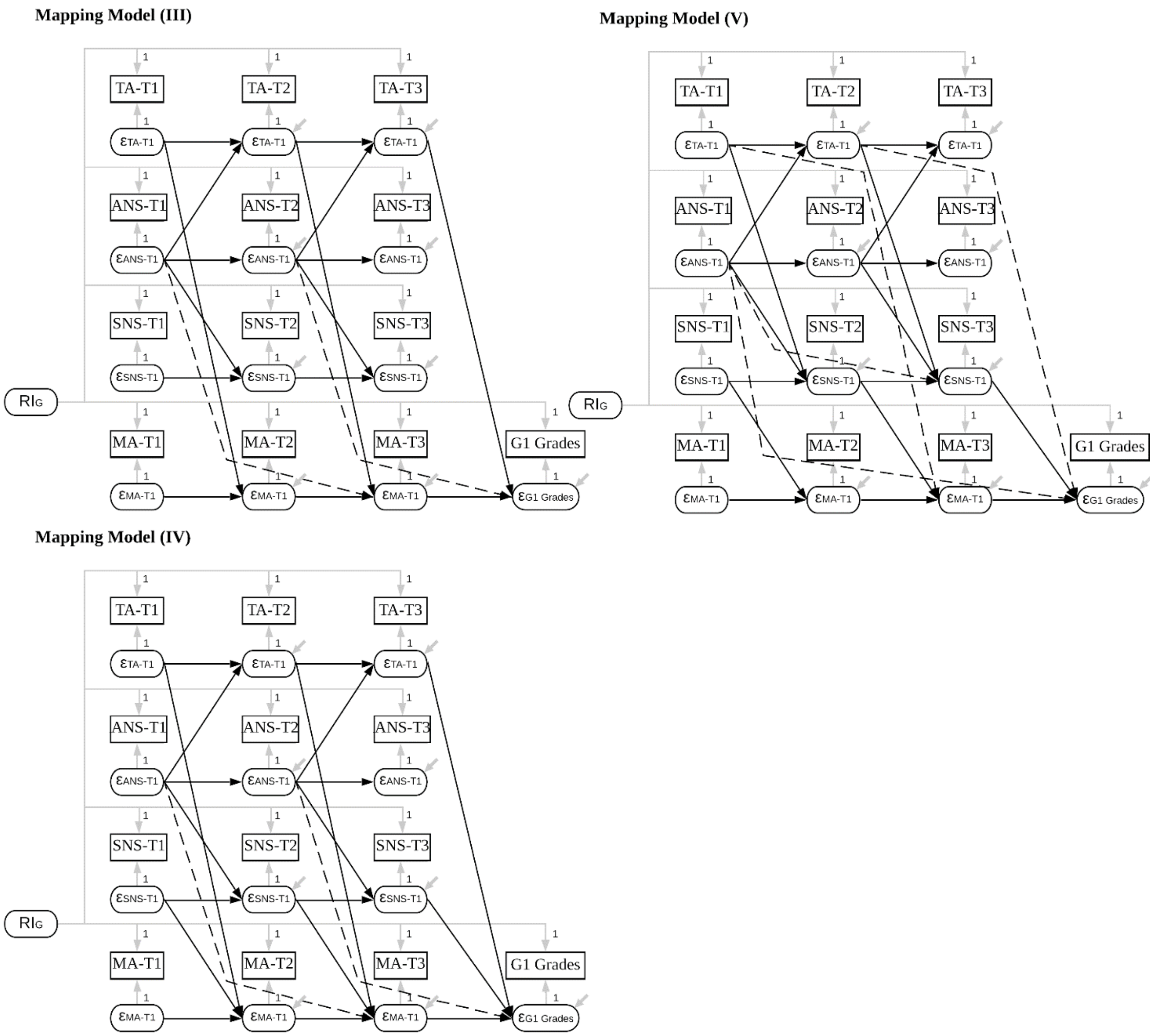
For Mapping Model (III), results indicate that model fit is near-acceptable, $\chi^2_{(182)} = 526.960$, $p < 0.001$, AIC = 50323.031, BIC = 50930.691, CFI = 0.949, TLI = 0.898, RMSEA = 0.055 [90% CI: 0.050, 0.061], SRMR = 0.050. However, Mapping Model (III) had inferior model fit when compared with the Time-Dependent/Bi-Directional Model, ΔAIC = 150.38, ΔBIC = 88.319, $\chi^2(14) = 178.38$, p < 0.001, and the Refinement Model, ΔAIC = 140.741, ΔBIC = 149.607, $\omega^2 = 0.601$, $p < 0.001$; z = 4.493, $p < 0.001$.

For Mapping Model (IV), results indicate that model fit is adequate, $\chi^2_{(179)} = 480.520$, $p < 0.001$, AIC = 50282.591, BIC = 52436.999, CFI = 0.955, TLI = 0.909, RMSEA = 0.052 [90% CI: 0.046, 0.058], SRMR = 0.047 However, Mapping Model (IV) had inferior model fit when compared with the Time-Dependent/Bi-Directional Model, ΔAIC = 109.94, ΔBIC = 61.177, $\chi^2(11) = 131.94$, p < 0.001, and the Refinement Model, ΔAIC = 100.301, ΔBIC = 122.465, $\omega^2 = 0.529$, $p < 0.001$; z = 3.652, $p < 0.001$.

For Mapping Model (V), results indicate that model fit is adequate, $\chi^2_{(181)} = 505.873$, $p < 0.001$, AIC = 50265.944, BIC = 50850.958, CFI = 0.955, TLI = 0.917, RMSEA = 0.050 [90% CI: 0.044, 0.055], SRMR = 0.051. However, Mapping Model (IV) had inferior model fit when compared with the Time-Dependent/Bi-Directional Model, ΔAIC = 114.987, ΔBIC = 57.359, $\omega^2 = 0.490$, $p < 0.001$; z = 4.934, $p < 0.001$, and the Refinement Model, ΔAIC = 105.348, ΔBIC = 118.647, $\omega^2 = 0.536$, $p < 0.001$; z = 3.981, $p < 0.001$.

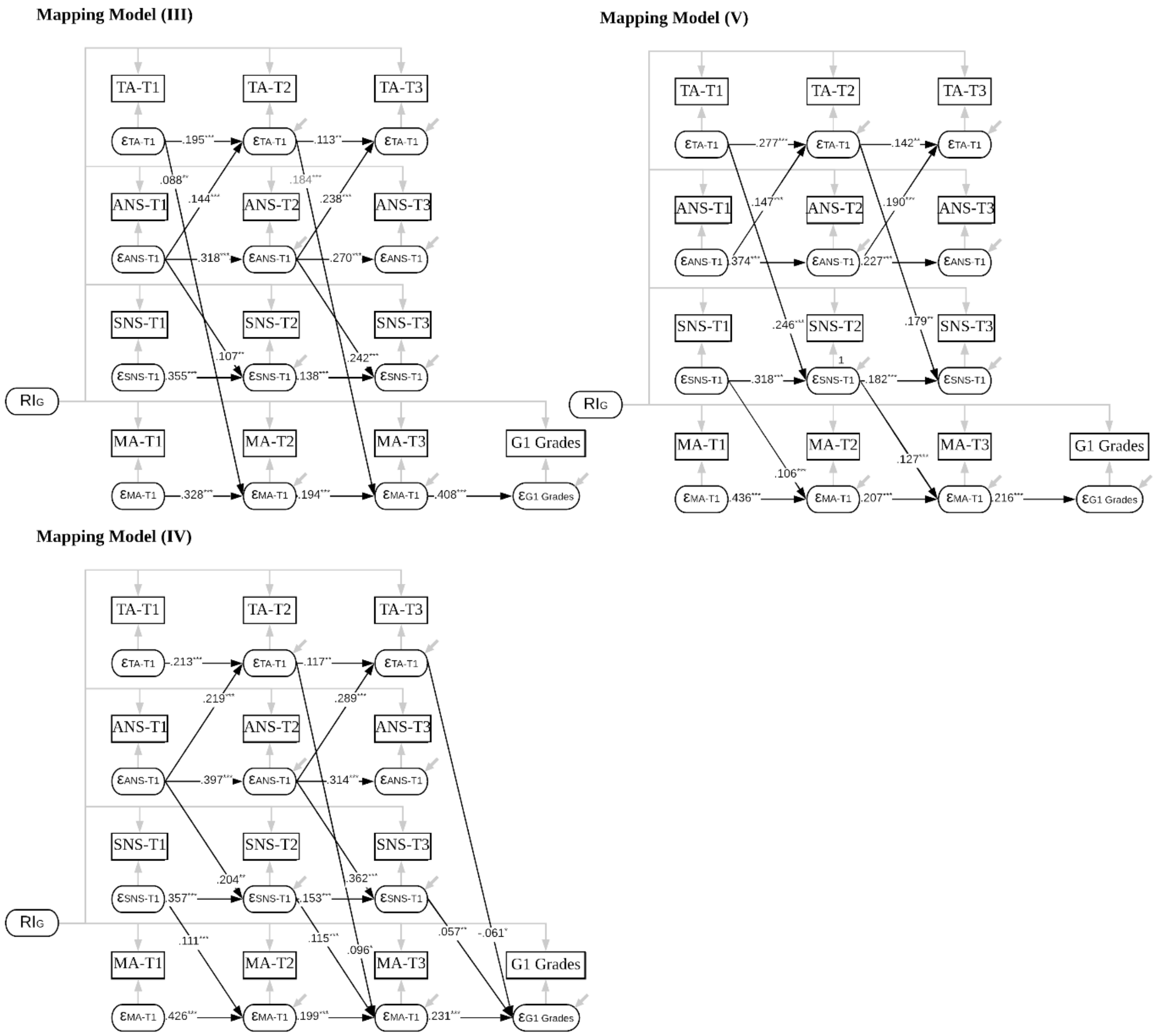
In sum, after considering Mapping Model (III - V), both Time-Dependent/Bi-Directional Model and Refinement Models remained the best fitting models (see **Figure 4**).

Figure 3 - ***Mapping Models (III - V)***



Note: TA refers to translating ability and MA refers to math achievement. T1-3 refers to time-points of data collection. Within each time-point, ANS, SNS, translating ability and math achievement residuals are allowed to co-vary. Note that no covariates (described above), residuals, or covariations are illustrated for the sake of clarity. Grey lines are specifications for the random intercepts and structured residuals. Black lines are the autoregressive and cross-lagged coefficients of interest.

Figure 4 - Standardized Coefficients for Mapping Models (III- V)



Note. TA refers to translating ability and MA refers to math achievement. T1-3 refers to time-points of data collection. Grey lines are specifications for the random intercepts and structured residuals. Black lines are the autoregressive and cross-lagged coefficients of interest. Statistically insignificant path coefficients are not shown for clarity. Grey lines are specifications for the random intercepts and structured residuals and are not numerated for clarity.

2.8 Discussion for Study 1

Through the estimation of multiple RI-CLPMs representing different accounts of development, study 1has directly compared whether, when between-individual differences are controlled for, the observed patterns of development support the mapping or refinement accounts. Three significant research findings were uncovered: 1) The random-intercept of the four main variables of interest (reflecting the underlying trait-like factors that underlie these variables) approached unity, 2) that model fit for all models tested were acceptable or approached acceptability, and 3) that the observed within-individual patterns of development best reflect the refinement account.

2.8.1 Overlapping of Trait-Like Individual Difference

In the formulation of the four RI-CLPMs, we have found that the random intercepts for the four main variables of interest approached unity. This suggests that there is substantial overlapping of the time-invariant trait-like individual differences that underlie these four constructs. That the individual and environmental factors underlying the four variables of interest are substantially overlapping is not unprecedented. For example, the correlation between the random intercepts of two traditionally distinct domains – math and reading – was found to be 0.94 (Bailey et al., 2020). Given that our four main variables measure different aspects of the same domain, and that the task configurations are very similar, it is not surprising that the random intercepts approached unity.

Theoretically, this is also quite plausible for both the mapping account and refinement account. Both accounts posit that the processing of symbolic and non-symbolic numbers is supported by overlapping neural mechanisms (e.g., Sokolowski et al., 2017). The difference between the two accounts is in how tightly linked symbolic and non-symbolic numbers are early in development and growth in which of the two drives growth.

Nevertheless, the substantial overlapping of the random intercepts calls into question just how separate these typically operationalized constructs are (at least at the between-person level). These results highlight the importance for future longitudinal and experimental studies to examine the empirical distinguishability of these constructs – that is, whether the conceptual differences between these variables have any practical significance. Furthermore, if indeed these concepts are distinguishable, it would be interesting to e whether there are differences in the predictiveness attributable to the shared developmental origins of these constructs versus the predictiveness attributable to factors unique to the construct of interest.

2.8.2 Implication of High Model Fit for all Tested Models

One of the most striking facts regarding our results is that all the tested models in the study have adequate or near adequate fit (see **Table 4**), with all or most of the hypothesized path coefficients within any hypothesized model being statistically significant (see **Figure 2**). This may be because the three comparison tasks share a large amount of common variance. While we did not detect multicollinearity in our sample (all VIFs < 5; standard errors of path coefficients are not abnormally large), the large amount of shared variance between the three variables allowed for relatively good model fit for any model when examined in isolation. Put another way, the shared variance of variables allowed for a certain amount of interchangeability in the independent variable of choice, and that a relatively poorly fitted model will still "adequately" fit the observed data in the absence of competing models.

This has two implications for the interpretation of existing studies and the practice of future studies. First, the high intercorrelation of the variables is problematic statistically because most classical methods of testing (e.g., multiple regression, mediation analysis, etc.) are guided by statistical significance. It is therefore easy to yield results supporting different theoretical accounts by changing which independent variables are hypothesized and included in analyses. This suggests that confirmatory evidence – that is, non-experimental evidence confirming specific hypotheses regarding development in numerical ability – would have limited utility. Instead, it may be more fruitful if future studies were to explore disconfirming evidence that could discriminate between different accounts of numerical development (e.g., through training studies of either symbolic or non-symbolic abilities).

This finding may contribute to reconciling the contradicting results regarding the predictive ability of ANS acuity in future math achievement (for a review, see Chen & Li, 2014). While some studies have found evidence for such a relation (e.g., Park & Brannon, 2013), other studies failed to find such a relation (e.g., Sasanguie et al., 2014). Our results suggest that a major potential reason for the inconsistent results may be due to the common practice of generating confirmatory evidence based on statistical significance to support a specific developmental account (in contrast to generating disconfirming evidence to discriminate between developmental accounts). As our results have shown, the high intercorrelations among the variables allows for relatively poorly fitting models to still be statistically significant. Therefore, apparently conflicting results in the extant literature may be an artifact of not directly contrasting different developmental models with one another. In light of these findings, researchers would be strongly encouraged to consider competing models when forming hypotheses on numerical development. Further, given that numerical development is ostensibly a within-individual phenomena, it would be fruitful for researchers to statistically control for between-person variability (i.e., effects from unobserved domain general confounds or stable environmental factors) when considering the developmental relations between different magnitude representation abilities. One possible method is employing a RI-CLPM approach as we have done in the current study.

2.8.3 Mapping Account and Refinement Account

The models tested in this study concerned the longitudinal predictive relations of ANS, SNS, translating ability, and math achievement. We aimed to contrast two theories of early numerical development by testing the theoretical predictions laid out by the two accounts. Using a longitudinal, repeated-measures design on a sample of 622 kindergarteners, we provided a comprehensive test of the time-lagged predictive relations between the ANS and SNS early in formal schooling.

Our analyses within the SEM framework allowed us to directly compare the relative fit of the implied variance-covariance matrix of the models derived from the different theoretical accounts against the observed variance-covariance matrix. This is particularly advantageous as individual paths within a hypothesized model may turn out significant, but the overall model would nevertheless be an inferior description of the causal relations between variables. As the use of the SEM framework provided measures of relative fit between different models over and beyond that of the significance of the individual paths, we were able to evaluate and contrast each model’s fit to the observed data. As such, our results suggest that the theoretical predictions of the refinement account not only adequately fit the observed data, but also provided a superior description of the observed data when compared to the theoretical predictions provided by the mapping account.

Through the main analyses, we found substantial evidence favoring the view that SNS abilities at the beginning of senior kindergarten significantly and consistently predict growth in later ANS and the ability to translate between the two systems (hypotheses IIa and IIb). Further, we also found evidence that is most consistent with the view that SNS abilities are consistently predictive of math achievement (hypothesis IIc). Finally, our results suggest that there is little evidence for the existence of a bi-directional relation between ANS and SNS. Importantly, our results suggest that these are within-person effects – that is, growth in SNS abilities within a child in an earlier time-point is predictive of later ANS abilities, translating ability, and math achievement within the same child.

Our results support the view that children's SNS and ANS are relatively separate at the beginning of development, and that growth in symbolic number processing facilitates the growth of later non-symbolic abilities (Mix et al., 2002; Noël & Rousselle, 2011). These results are in line with previous longitudinal studies that link changes in earlier SNS acuity with later ANS acuity in kindergarteners and first graders (Matejko & Ansari, 2016; Mussolin et al., 2014). Further, our findings are also concordant with previous research suggesting that symbolic number knowledge, but not non-symbolic number knowledge, is predictive of math achievement (e.g., Holloway & Ansari, 2009; Sasanguie, Defever, Maertens, & Reynvoet, 2014; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013).

2.8.4 Limitations

A limitation of study 1 is that the tasks that are operationalized to measure ANS acuity, SNS acuity, translating ability, and math achievement are not typical. First, our comparison tasks are pencil-and-paper based and between-subject differences are calculated as the difference between the number of correct and incorrect trials. The presentation of multiple comparison tasks per page is psychometrically more limited as it presents students with noisier stimuli, and the use of pencil-and-paper is more impacted by differences in participants' hand-eye coordination, which may vary in kindergarten. Similarly, math achievement is operationalized as arithmetic verification and school grades. In addition to the limitations stated above, arithmetic verification is limited as it does not require students to come up with the answer and the measure is somewhat narrower in scope when compared with typically measures of math achievement. Finally, school grades are limited as it not a pure measurement of math achievement – as it may also include a teacher's evaluation of student effort and good behavior.

It is important to note, however, that the Numeracy Screener (Nosworthy et al., 2013) used in study 1 had yielded comparable results in multiple studies across different cultural contexts and age groups (Fischer et al., 2020; Hawes et al., 2019; Soltani & Mirhosseini, 2019; Susperreguy et al., 2020; Tavakoli, 2016). This suggests that while the tasks employed in the numeracy screener may be different from more conventional tasks, the measurement of the underlying construct is preserved. Further, the problem of the elicitation of domain-general cognitive resources is not a problem solely found in study 1. Indeed, multiple studies have found that different ways of controlling for continuous properties of the dot arrays used in non-symbolic comparison tasks elicit different domain-general cognitive resources (e.g., Gilmore et al., 2014; Price et al., 2012), and there is currently substantial debate as to whether it is possible to garner a "pure" measurement of ANS acuity. Given that extant studies suggest that the numeracy screener yields comparable results, it is advantageous to use the Numeracy Screener to collect data from hard-to-reach populations.

2.9 Conclusion

In sum, using a three time-point longitudinal design, we tested four models of development informed by existing theories of mathematical development, specifically the theoretical predictions offered by the mapping account and the refinement account. Results suggested that the refinement account to best describe the developmental trajectory of students, and the relations between variables posited by the refinement account are consistent over multiple time-points sampled from the beginning of senior kindergarten to the end of grade 1. Further, results also speak against the notion that the ANS and SNS have a bi-directional relation.

2.10 References

Agrillo, C. (2015). Numerical and arithmetic abilities in non-primate species. R. Cohen Cadosh & Ann Dowker (Eds.), The Oxford Handbook of Numerical Cognition, 214–236. https://doi.org/10.1093/oxfordhb/9780199642342.013.002

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. Organizational Research Methods, 16(2), 270–301. https://doi.org/10.1177/1094428112470848

Becker, C., & Gather, U. (1999). The masking breakdown point of multivariate outlier identification rules. Journal of the American Statistical Association, 94(447), 947–955. https://doi.org/10.1080/01621459.1999.10474199

Berry, D., & Willoughby, M. T. (2017). On the practical interpretability of cross‐lagged panel models: Rethinking a developmental workhorse. Child Development, 88(4), 1186−1206.

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(4), e93565. https://doi.org/10.1371/journal.pone.0093565

Carey, S. (2004). Bootstrapping & the origin of concepts. Daedalus, 133(1), 59–68. https://doi.org/10.1162/001152604772746701

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. https://doi.org/10.3758/bf03196206

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. Journal of Family Psychology, 19(2), 294. https://doi.org/10.1037/0893-3200.19.2.294

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

De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. Journal of Experimental Child Psychology, 103(2), 186–201. https://doi.org/10.1016/j.jecp.2009.01.004

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(2), 48–55. https://doi.org/10.1016/j.tine.2013.06.001

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

De Vos, T. (1992). Tempo-Test-Rekenen. Handleiding.[Tempo Test Arithmetic. Manual]. Nijmegen: Berkhout.

Dehaene, S. (1999). The Number Sense: How the Mind Creates Mathematics. Oxford University Press.

Dehaene, S. (2005). Evolution of human cortical circuits for reading and arithmetic: The "neuronal recycling" hypothesis. From Monkey Brain to Human Brain, 133–157.

Dehaene, S. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In Sensorimotor foundations of higher cognition (pp. 527–574).

DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: Effects of feedback and training. Frontiers in Human Neuroscience, 6. https://doi.org/10.3389/fnhum.2012.00068

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

Enders, C. K. (2001). A Primer on Maximum Likelihood Algorithms Available for Use With Missing Data. Structural Equation Modeling: A Multidisciplinary Journal, 8(1), 128–141. https://doi.org/10.1207/S15328007SEM0801_7

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

Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. Child Development Perspectives, 7(2), 74–79. https://doi.org/10.1111/cdep.12019

Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. Current Directions in Psychological Science, 22(1), 23–27.

Giles, P. (2004). Low income measurement in Canada. Statistics Canada Ottawa.

Gimbert, F., Camos, V., Gentaz, E., & Mazens, K. (2019). What predicts mathematics achievement? Developmental change in 5-and 7-year-old children. Journal of Experimental Child Psychology, 178, 104–120.

Goodman, J. S., & Blum, T. C. (1996). Assessing the non-random sampling effects of subject attrition in longitudinal research. Journal of Management, 22(4), 627–652.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the ‘number sense’: The approximate number system. In 3-, 4-, 5-, 6-Year-Olds and Adults. Developmental Psychology, 1457–1465. https://doi.org/10.1037/a0012682

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. Psychological Methods, 20(1), 102. https://doi.org/10.1037/a0038889

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1–55. https://doi.org/10.1080/10705519909540118

Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. Cognition, 131(1), 92–107. https://doi.org/10.1016/j.cognition.2013.12.007

Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. Proceedings of the National Academy of Sciences, 106(25), 10382–10385. https://doi.org/10.1073/pnas.0812142106

Jang, S., & Cho, S. (2018). The mediating role of number-to-magnitude mapping precision in the relationship between approximate number sense and math achievement depends on the domain of mathematics and age. Learning and Individual Differences, 64, 113–124. https://doi.org/10.1016/j.lindif.2018.05.005

Janus, M., Brinkman, S., Duku, E., Hertzman, C., Santos, R., Sayers, M., Schroeder, J., & Walsh, C. (2007). The early development instrument: A population-based measure for communities. A Handbook on Development, Properties, and Use. Hamilton, ON: Offord Centre for Child Studies.

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first‐grade math achievement from developmental number sense trajectories. Learning Disabilities Research & Practice, 22(1), 36–46. https://doi.org/10.1111/j.1540-5826.2007.00229.x

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. Developmental Psychology, 45(3), 850. https://doi.org/10.1037/a0014939

Lau, N. T., Merkley, R., Tremblay, P., Zhang, S., De Jesus, S., & Ansari, D. (2021). Kindergarteners' symbolic number abilities predict nonsymbolic number abilities and math achievement in grade 1. Developmental Psychology.

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. https://doi.org/10.1016/j.cognition.2006.10.005

Libertus, M. E. (2015). The role of intuitive approximation skills for school math abilities. Mind, Brain, and Education, 9(2), 112–120. https://doi.org/10.1111/mbe.12072

Libertus, M. E., Feigenson, L., Halberda, J., & Landau, B. (2014). Understanding the mapping between numerical approximation and number words: Evidence from Williams syndrome and typical development. Developmental Science, 17(6), 905–919.

Little, T. D. (2013). Longitudinal structural equation modeling. Guilford press.

Lonnemann, J., Linkersdörfer, J., Hasselhorn, M., & Lindberg, S. (2011). Symbolic and non-symbolic distance effects in children and their connection with arithmetic skills. Journal of Neurolinguistics, 24(5), 583–591. https://doi.org/10.1016/j.jneuroling.2011.02.004

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(4), 635. https://doi.org/10.1037/a0027248

Lyons, I. M., Bugden, S., Zheng, S., De Jesus, S., & Ansari, D. (2018). Symbolic number skills predict growth in nonsymbolic number skills in kindergarteners. Developmental Psychology, 54(3), 440. https://doi.org/10.1037/dev0000445

Martin, M. A., & Roberts, S. (2010). Jackknife-after-bootstrap regression influence diagnostics. Journal of Nonparametric Statistics, 22(2), 257–269. https://doi.org/10.1080/10485250903287906

Matejko, A. A., & Ansari, D. (2016). Trajectories of symbolic and nonsymbolic magnitude processing in the first year of formal schooling. PloS One, 11(3), e0149863. https://doi.org/10.1371/journal.pone.0149863

McCrink, K., & Wynn, K. (2009). Operational momentum in large-number addition and subtraction by 9-month-olds. Journal of Experimental Child Psychology, 103(4), 400–408. https://doi.org/10.1016/j.jecp.2009.01.013

Merkle, E. C., You, D., & Preacher, K. J. (2014). Testing non-nested structural equation models. ArXiv:1402.6720 [Stat]. http://arxiv.org/abs/1402.6720

Mix, K. S. (2008). Surface similarity and label knowledge impact early numerical comparisons. British Journal of Developmental Psychology, 26(1), 13–32. https://doi.org/10.1348/026151007X189109

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). Quantitative development in infancy and early childhood. Oxford University Press.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. Nature, 215(5109), 1519.

Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. Journal of Experimental Child Psychology, 103(4), 490–502. https://doi.org/10.1016/j.jecp.2009.02.003

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. https://doi.org/10.1371/journal.pone.0091839

Mussolin, C., Nys, J., Leybaert, J., & Content, A. (2012). Relationships between approximate number system acuity and early symbolic number abilities. Trends in Neuroscience and Education, 1(1), 21–31.

Mussolin, C., Nys, J., Leybaert, J., & Content, A. (2016). How approximate and exact number skills are related to each other across development: A review. Developmental Review, 39, 1–15. https://doi.org/10.1016/j.dr.2014.11.001

Muthén, L. K., & Muthén, B. O. (2010). Mplus: Statistical analysis with latent variables: User's guide. Muthén & Muthén Los Angeles.

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

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. https://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. https://doi.org/10.1016/j.cognition.2015.01.008

Ontario Ministry of Education and Training. (2005). The Ontario Curriculum, Grades 1-8: Mathematics. Revised. The Ministry.

Paglin, M., & Rufolo, A. M. (1990). Heterogeneous human capital, occupational choice, and male-female earnings differences. Journal of Labor Economics, 8(1, Part 1), 123–144. https://doi.org/10.1086/298239

Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. Psychological Science, 24(10), 2013–2019. https://doi.org/10.1177/0956797613482944

Parsons, S., & Bynner, J. (2005). Does numeracy matter more?

Pinheiro-Chagas, P., Wood, G., Knops, A., Krinzinger, H., Lonnemann, J., Starling-Alves, I., Willmes, K., & Haase, V. G. (2014). In how many ways is the approximate number system associated with exact calculation? PloS One, 9(11), e111155. https://doi.org/10.1371/journal.pone.0111155

Price, G. R., & Fuchs, L. S. (2016). The Mediating Relation between Symbolic and Nonsymbolic Foundations of Math Competence. PLOS ONE, 11(2), e0148981. https://doi.org/10.1371/journal.pone.0148981

Purpura, D. J., Baroody, A. J., & Lonigan, C. J. (2013). The transition from informal to formal mathematical knowledge: Mediation by numeral knowledge. Journal of Educational Psychology, 105(2), 453. https://doi.org/10.1037/a0031753

Rivera-Batiz, F. L. (1992). Quantitative literacy and the likelihood of employment among young adults in the United States. Journal of Human Resources, 313–328. https://doi.org/10.2307/145737

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). Journal of Statistical Software, 48(2), 1–36.

Rousselle, L., & Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. Cognition, 102(3), 361–395. https://doi.org/10.1016/j.cognition.2006.01.005

Rowley, G. L., & Traub, R. E. (1977). Formula Scoring, Number-Right Scoring, and Test-Taking Strategy. Journal of Educational Measurement, 14(1), 15–22.

Smith, F. M., & Hausafus, C. O. (1998). Relationship of family support and ethnic minority students' achievement in science and mathematics. Science Education, 82(1), 111–125. https://doi.org/10.1002/(SICI)1098-237X(199801)82:1<111::AID-SCE6>3.0.CO;2-K

Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: A neural model. Journal of Cognitive Neuroscience, 16(9), 1493–1504. https://doi.org/10.1162/0898929042568497

Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. Econometrica, 57(2), 307–333. https://doi.org/10.2307/1912557

Wong, T. T.-Y., Ho, C. S.-H., & Tang, J. (2016). The relation between ANS and symbolic arithmetic skills: The mediating role of number-numerosity mappings. Contemporary Educational Psychology, 46, 208–217. https://doi.org/10.1016/j.cedpsych.2016.06.003

Woodcock, R. W., Johnson, M. B., & Mather, N. (1990). Woodcock-Johnson psycho-educational battery—Revised. DLM Teaching Resources.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. Cognitive Psychology, 24(2), 220–251. https://doi.org/10.1016/0010-0285(92)90008-P

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. Cognition, 74(1), B1–B11.

Zhang, J., & Norman, D. A. (1995). The representation of numbers. Cognition, 57, 271–295.

Chapter 3

3.1 Introduction

Proficiency in mathematics is an important predictor of income, career success and overall well-being (Paglin & Rufolo, 1990; Parsons & Bynner, 2005b; Rivera-Batiz, 1992). In recent decades, the ability to represent quantities with non-iconic symbols (e.g., spoken numbers, Arabic numbers, etc.) has emerged as a foundational skill that allows for the acquisition of other early numerical skills, and had been hypothesized to be a predictor of children's later math achievement (Feigenson et al., 2004). In recent years, two developmental accounts have emerged to describe how children initially acquires the ability to represent quantities with symbols, and there currently exist substantial debate regarding which of the two developmental accounts more accurately describes children's development. To begin, we shall describe the two developmental accounts.

3.1.1 The Mapping Account

Perhaps the most prominent of the two accounts is the so-called mapping account. Under the mapping account, when children are first learning to represent quantities with symbols, they connect symbols with a pre-existing system for the approximate processing and representation of sets of items (often referred to as the approximate number system (ANS); Dehaene, 1999; Feigenson, Dehaene, & Spelke, 2004). It is argued that the initially meaningless arbitrary symbols acquire meaning by a one-to-one mapping of the symbol onto the corresponding quantity of the ANS (Dehaene, 2007; Piazza, 2011). Consequently, ANS ability, often assessed by the ability to compare non-symbolic quantities (e.g., •• vs. ••••••), and symbolic number ability, often assessed by the ability to compare symbolic quantities (e.g., 2 vs. 6), are hypothesized to draw from the same underlying representation, and are developmentally tightly linked.

Consistent with this mapping account, multiple studies have found evidence to suggest that earlier variability in ANS ability is predictive of later math achievement (Bonny & Lourenco, 2013; Libertus et al., 2011, 2013; Mazzocco et al., 2011; Mundy & Gilmore, 2009). Some studies suggest that this relation between ANS ability and math achievement is a direct relation – just as symbolic numbers and ANS share the same neural underpinnings, the manipulation of symbolic numbers would similarly be reliant on the neural underpinnings that allow for non-symbolic number manipulation (Dehaene, 2005; Feigenson et al., 2004; Gilmore et al., 2010). Other studies suggest that the relation may be indirect – with ANS ability predicting symbolic number ability, which in turn, predict later math achievement (Price & Fuchs, 2016).

In recent years, however, studies suggest that the relation between ANS ability and math achievement may not be straightforward. In contrast to previous findings and the theoretical predictions of the mapping account, a number of studies have reported no significant relation between ANS ability and math achievement (e.g. Ferreira et al., 2012; Fuhs & McNeil, 2013; Kolkman et al., 2013; Sasanguie, De Smedt, et al., 2012; Sasanguie, Van den Bussche, et al., 2012).

In this context, it has been suggested that differences in methods of experimentally controlling for continuous properties (e.g., total area, individual dot sizes, density, etc.) of dot arrays may be the culprit of these inconsistent results. Specifically, it has been suggested that the different ways researchers control for continuous properties of dot arrays may elicit the need for participants to inhibit incongruent information. For example, the task of comparing three large dots with five small dots (in terms of quantity) would require participants to supress the incongruent information of dot size (because there exists a conflict between quantity and size) and focus on the numerosity of the arrays. In contrast, the comparison of three small dots against five large dots would require no such inhibition. It is therefore proposed that non-symbolic comparison tasks do not simply tap into numerical magnitude processing but instead also engage participants' inhibitory control. Indeed it has been shown that inhibitory control explains individual differences in non-symbolic number comparison and explains part of the variance of the relationship between ANS tasks and measures of children's numerical and mathematical abilities (Clayton & Gilmore, 2015; Gilmore et al., 2013; Szucs et al., 2013).

3.1.2 The Refinement Account

In light of these controversial findings, the refinement account has been proposed as a potentially more accurate account of numerical development. Under the refinement account, the central claim of the mapping account – that the learning of numerical symbols involves a mapping between the symbol and some innate number ability – is only an intermediary step for a full understanding of symbolic numbers to emerge.

Under the refinement account, an initial and limited semantic understanding of symbolic numbers is acquired through a one-to-one mapping of numbers 1-4 onto the object tracking system (Carey, 2001, 2004), an innate visual system that allows for the perception and tracking of up to four objects (Scholl, 2001; Simon, 1997). This initial understanding is combined with the learning of counting lists at home or in preschool to infer characteristics of the number system (i.e., order, cardinality, and the successor function). Gradually, these learned characteristics are applied to numbers larger than 4, and enables an initial and basic understanding of the symbolic number system (Carey, 2001, 2004, 2009).

Once this basic understanding of the symbolic number system emerges, the repeated practice and usage of symbolic numbers leads to a gradual shift in the sematic content of the numerical symbols. Numerical symbols initially based on the object tracking system would be increasingly overwritten and reinterpreted by their relations with other numerical symbols (Mix et al., 2002; Nieder, 2009).

3.1.3 Two Outstanding Issues

In study 1, we employed several RI-CLPM models to ascertain which of these two developmental accounts best fit developmental data, and results have largely supported the refinement account as the better fitting model. Study 2 compliments the results from study 1 by examining the degree to which all members in the population conform to the developmental trajectory as hypothesized by the refinement account. In other words, study 2 aims to ascertain whether either the mapping or refinement accounts adequately describes the developmental trajectories of the whole population (as opposed to only part of the population).

3.1.3.1 Distinct Subgroups with Different Developmental Profiles. One seldomly examined issue in the extant literature pertains to whether the population is homogenous or heterogenous. Specifically, several research studies have found evidence to suggest that there are distinct subgroups in the population that may exhibit different patterns of numerical development (Bartelet et al., 2014; Chew et al., 2016, 2019; de Souza Salvador et al., 2019; Kißler et al., 2021).

For example, using latent class analysis, Chew et al. (2016) examined patterns of performance in children's symbolic comparison and non-symbolic comparison accuracy and reaction time. These authors found two distinct groups of children with comparable symbolic comparison accuracy scores, but one group had significantly slower reaction time than the other. The authors posited that the relatively slower group may be completing symbolic comparison task differently due to possessing weaker one-to-one associations between symbolic numerals and the corresponding non-symbolic representation in the ANS. According to this account, there exists a subgroup of children who have weaker mapping between ANS and symbolic numbers and therefore likely employ more inefficient strategies to complete the task (Chew et al., 2016).

These findings are problematic for the abovementioned developmental accounts and the associated debates. This is because both developmental accounts posit one direction of developmental relationships between ANS and symbolic numbers. Given the findings of different profiles of developmental change, it may be possible that different subgroups fit different models of the relationship between ANS and symbolic numbers. For example, under the mapping account, children acquire symbolic number ability by affixing the symbols onto the corresponding quantity in the ANS (Dehaene, 2007; Piazza, 2011). The finding that children use different strategies when completing the symbolic comparison task, and the implication that subgroups in the population may have varying strength of association between ANS ability and symbolic number ability would be contrary evidence against this idea that there is one underlying mechanism of numerical development. Instead, these findings suggest that the developmental accounts may be incomplete, and further suggest that multiple mechanisms may be required to properly describe how different subgroups may acquire symbolic number knowledge.

Importantly, the existence of subgroups may be a potential reason for the observed divergence in research findings. If groups of children exhibit different patterns of association between ANS ability, symbolic number ability and math achievement, then studies of the population ignoring these subgroups would yield parameter estimates that reflects an aggregation these different developmental patterns (Hickendorff et al., 2018). The observed divergence in research findings could then be due to different proportions of subgroups that are the consequences of sampling. For example, suppose one subgroup of the population exhibit a strong relation between ANS ability and math achievement, while another subgroup exhibits no such relation. A research project that happens to sample more children from the first subgroup may find significant associations between the variables, while a research project that happens to sample more children from the second subgroup may fail to find a relation between the variables.

In sum, the mismatch between the assumption of homogeneity in the theoretical accounts and the observed heterogeneity in populations is problematic and may have contributed to the divergence in research findings.

3.1.3.2 Within- and Between-Individual Variability. Similar to Study 1, concerns regarding within-individual and between-individual variability needs to be addressed to properly ascertain whether there are subgroups in the population with different developmental profiles.

Specifically, there currently is a mismatch between, on the one hand, the theoretical predictions of both developmental accounts, and on the other, the statistical methods used to verify these predictions. For example, while the mapping account posits the relation between ANS ability with symbolic number ability and math achievement as within-person processes (i.e., growth in a child's ANS ability is predictive of the same child's symbolic number ability and math achievement), research methods to verify these hypotheses often provide support for between-person associations (e.g., children with ANS ability tends to also be children who have stronger symbolic number ability and math achievement; Curran & Bauer, 2011; Hoffman & Stawski, 2009).

Critically, using between-person associations as evidence for within-person associations is appropriate only if it is assumed that between-person relation perfectly reflects an aggregation of within-person processes (Diez-Roux, 1998). This assumption is often not correct and between-person associations can often differ significantly in size and polarity from the within-person associations. An example from the medical literature can demonstrate this point. It has been demonstrated an individual have a higher probability of having a heart attack if he or she is exercising (i.e., a within-person association). However, simultaneously, people who tends to exercise more are people who also tend to have a lower risk of heart attack (i.e., the between-individual effect; Curfman, 1993).

Importantly, this limitation is applicable to extant studies pointing to the possibility of subgroups in the population (e.g., Bartelet et al., 2014; Chew et al., 2016, 2019; de Souza Salvador et al., 2019; Kißler et al., 2021), and the identified subgroups in these studies are generated from a mixture of within-person and between-person variability. This makes the interpretation of these subgroups problematic, as it is unclear whether the grouping reflect states of development or trait differences between subgroups.

3.1.4 Study 2

In light of these limitations, study 2 bridges this research gap by employing a random-intercept latent transition analysis (RI-LTA; Muthén & Asparouhov, 2020). Importantly, the RI- partitions out of between-person influences, ascertains whether there are subgroups of the population with different within-person, state-like patterns of development, and examines how different subgroups may transition between groups over time.

Three research questions will be explored: 1) whether when between-person differences are accounted for, there are subgroups in the population that exhibits different developmental profiles, 2) whether the patterns of transitions suggest that there are subgroups in the population, and 3) whether membership in different profiles predict children's math achievement.

3.2 Methods

Study 2 utilizes a subset of the dataset used in study 1, this subset was combined with additionally collected data regarding students' academic performance. The dataset came from an ongoing collaboration between the University of Western Ontario (UWO) and the Toronto District School Board (TDSB). The ongoing collaboration between the TDSB and UWO was approved by the TDSB's External Research Review Board (ERRC), and testing material was approved by the University of Western Ontario's Non-Medical Research Ethics Board (REB#102158: "The relationship between symbolic and non-symbolic numerical magnitude processing and arithmetic achievement in primary school"). Sensitive information regarding parents and students was kept confidential within the TDSB's Research and Information Services and only anonymized data were shared with researchers at UWO.

3.2.1 Participants

Participants came from TDSB schools in the greater Toronto area. Students were assessed at 62 months, 68 months and 80 months in age (SD = 3.53), which corresponded to the beginning of senior kindergarten (T1), six-months into senior kindergarten (T2), and six-months into Grade 1 (T3). The final sample consisted of 622 children (279 females).

3.2.2 Experimental Tasks

Experimental tasks were administered using paper-and-pencil numeracy screener booklets (https://osf.io/xgk89/), based on the designs by Nosworthy, Bugden, Archibald, Evans, & Ansari (2013). At each time point, children completed the symbolic number comparison task, non-symbolic number comparison task and mixed comparison task – always in that order. Each comparison task consisted of 72 total items with 12 items per page. Children were given 2 minutes to complete as many items as quickly and accurately as possible. Further, they were instructed to not skip any items. The same numeracy booklet was used for all three timepoints.

3.2.3 Scoring

For all three experimental tasks, raw scores were the total number of items completed within the two-minute time limit. To adjust for guessing, we followed Rowley & Traub's (1977) recommendation and submitted the raw score to the formula (A = C – I), where A is the adjusted score, C is the total number of correct items, and I is the total number of incorrect items. For replication purposes, this scoring method is identical to Lyons et al. (2018).

3.2.4 Numeral Comparison Task

Children's symbolic number ability was measured through performance on the numeral comparison task. In each trial of the number comparison task, children were instructed to compare two Arabic numerals arranged side-by-side and to decide which of the two numbers was numerically larger. Numbers ranged from 1-9 with the absolute distance between the two numbers (i.e., $|n_1 – n_2|$) ranging from 1 to 3. In total, 18 possible pairs of numbers were included: all 15 number pairs with distances of 1 or 2, along with 3 number pairs of distance 3 (1_4, 3_6, and 6_9). The 18 possible pairs were arranged pseudo-randomly over four blocks, giving a grand total of 72 trials. The first block of 18 pairs was arranged so that 9 pairs had the larger numeral was on the left side, and the other 9 pairs had the larger numeral on the right side. The 9 trials were chosen such that the larger side was not related to numerical size, distance, or ratio. The next 18 were arranged in the opposite manner. The last 36 trials were determined in the same manner as the first two blocks. Trial order within each block of 18 such that, for an n^{th} item in the sequence, the average numerical ratio, size, and distance were the same across the three comparison tasks (numeral, dot, and mixed). This arrangement was done so that the size and distance ratios encountered on each task would not have significantly differed across tasks (all *p*s >.20), and thus not confound with the stimulus format.

3.2.5 Dot Comparison Task

Children's ANS was measured through performance on the dot comparison task. In each trial of the dot comparison task, children were instructed to compare which of two arrays of dots arranged side-by-side and decide which of the two arrays had more dots. The children were also instructed to use their best guess and not to count the dots. Numerosities and trial order were the same as the numeral comparison task. Additionally, two versions of the task were given. In one version, dot area positively correlated with numerosity, and overall contour length was negatively correlated with numerosity, the opposite was true in the other version. As such, relying on any single parameter would have led to chance performance. Version order was also pseudo-randomized so that it was not informative of the correct answer within a given segment of trials.

3.2.6 Mixed Comparison Task

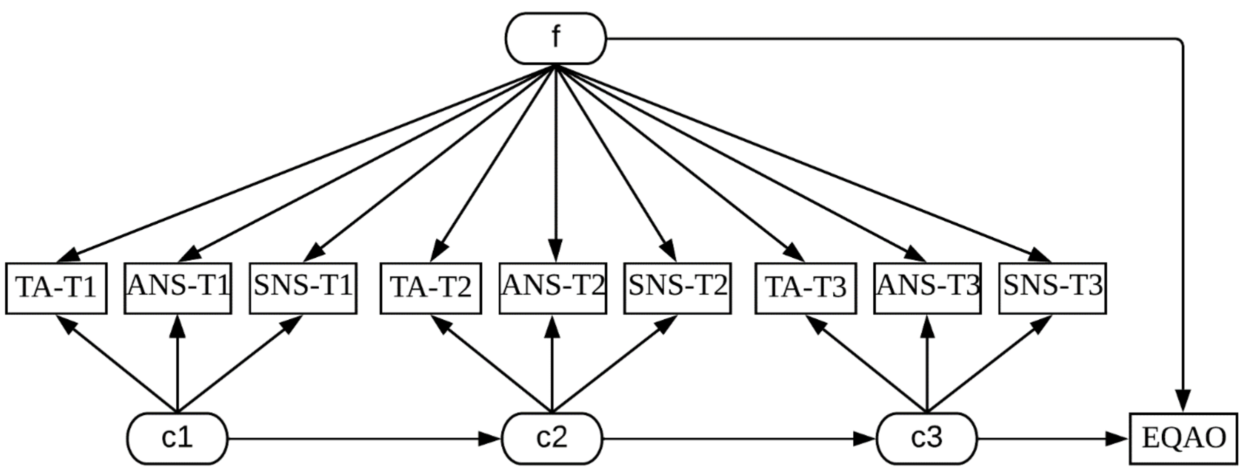
Children's translating ability was measured through performance on the mixed comparison task. In each trial of the mixed comparison task, children were instructed to compare an array of dots against an Arabic numeral arranged side-by-side and decide which of the two were numerically larger. The children were also instructed to use their best guess and not to count the dots. Numerosities and trial order were the same as the numeral comparison task. Further, the side which contained the numeral and which the dots were pseudo-randomized so that it was not informative of the correct answer.

3.2.7 Educational Quality and Accountability Office Assessment Score

Children's math ability at the end of grade 3 was assessed using the Educational Quality and Accountability Office (EQAO) Grade 3 score. The EQAO assesses elementary students at the end of grade 3 and grade 6 in reading, writing and mathematics. The mathematics EQAO assessment is composed of 25 questions administered to students by their teacher. Four levels of achievement are possible: at level 1, the student has demonstrated limited mastery of the required knowledge and skills to achieve provincial standards, at level 2, the student has demonstrated the required knowledge and skills that approaches provincial standards, at level3, the student the knowledge and skills to reach provincial standards, and at level 4, the student has demonstrated skills that exceed the provincial standard.

3.3 Analytic Procedures

Figure 5**.** Path diagram of RI-LTA.



Note. T1, T2 and T3 refers to time 1, time 2 and time 3, respectively. ANS refers to ANS ability, TA refers to translating ability, SNS refers to symbolic number ability, and EQAO refers to Grade 3 EQAO scores.

To ascertain whether there is evidence for subgroups in the population, we estimated a random-intercept latent transition analysis (RI-LTA; Muthén & Asparouhov, 2020) model. Please see **Figure 5** for an illustration of the model. The RI-LTA contains three features. First, the model contains a set of latent class variables (c1-c3 in **Figure 5**) that posits that an underlying, unobserved grouping variable can be inferred from a set of observed indicator variables (Collins & Lanza, 2009). Second, the RI-LTA allows for members in one group to transition between profile membership over time (as illustrated by the arrow pointing from c2 to c3, and c1 to c2 in **Figure 5**). Finally, the model contains a random intercept (f in **Figure 5**) whose factor loadings capture individual differences.

A model building approach was used to estimate the RI-LTA. First, to determine the optimal number of latent profiles, six exploratory RI-LTAs (from a two-profile to seven-profile model) was conducted. These models were assessed using fit statistics and substantive theory (described below). Second, the RI-LTA with the optimal number of profiles was examined. Finally, EQAO scores were added as a distal outcome to ascertain whether profile membership at grade 1 predicts EQAO scores at grade 3.

Table 5. Descriptive Statistics of Variables in Study 2

Variable	**Mean**	**SD**	**Skew**	**Kurtosis**
Number Comparison T1	19.10	15.33	0.68	0.30
Number Comparison T2	30.85	15.52	0.01	0.04
Number Comparison T3	45.98	14.57	-0.44	0.36
Dot Comparison T1	15.06	9.29	0.40	1.81
Dot Comparison T2	22.24	10.50	0.72	2.36
Dot Comparison T3	30.51	10.99	0.33	0.57
Mixed Comparison T1	12.96	10.28	0.36	0.25
Mixed Comparison T2	21.05	11.45	0.69	1.78
Mixed Comparison T3	31.15	11.47	0.21	0.85
EQAO scores	3.26	0.675	-0.46	-0.01

Model fit statistics and substantive theory are combined to decide on the best fitting model. Regarding model fit statistic, the Bayesian Information Criterion (BIC), adjusted Bayesian Information Criterion (ABIC) and entropy (ranging from 0 to 1) was used (Peel & McLachlan, 2000). Lower values for BIC and ABIC and higher values for entropy indicate better model fit. Models were estimated with Mplus 8.4 (Muthén & Muthén, 2010) using the maximum likelihood estimator. To avoid local maximums, all RI-LTAs were run with 4000 start values with 1000 iterations and retained the 200 best solutions for final stage optimization (Hipp & Bauer, 2006; Peel & McLachlan, 2000).

3.4 Results

3.4.1 Descriptive Statistics

Table 5 and **Table 6** descriptive statistics and the variance-covariance matrix, respectively. Examination of skewness and kurtosis revealed that all variables modestly deviated from normality (Skewness < 2, Kurtosis < 7; Finch, West, & MacKinnon, 1997).

Table 6. Correlation and Covariance Matrix for the Variables of Interest

Variables	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Number Comparison T1	**190.1**	0.72	0.48	0.73	0.61	0.51	0.77	0.62	0.46	0.52
2. Number Comparison T2	143.2	**210.9**	0.58	0.55	0.73	0.54	0.62	0.79	0.53	0.49
3. Number Comparison T3	96.2	121.9	**210.0**	0.40	0.51	0.76	0.47	0.51	0.74	0.49
4. Dot Comparison T1	84.0	67.0	48.1	**70.1**	0.59	0.46	0.71	0.53	0.39	0.44
5. Dot Comparison T2	78.0	98.1	68.5	46.1	**86.4**	0.51	0.55	0.74	0.50	0.44
6. Dot Comparison T3	74.9	84.9	117.6	41.1	50.6	**115.4**	0.51	0.55	0.74	0.50
7. Mixed Comparison T1	97.8	83.4	62.1	55.0	47.2	50.6	**84.8**	0.61	0.46	0.52
8. Mixed Comparison T2	87.9	117.4	76.6	45.9	71.1	55.6	57.7	**106.0**	0.52	0.47
9. Mixed Comparison T3	70.4	86.1	119.1	36.0	51.7	88.8	46.8	60.3	**123.9**	0.49
10. EQAO Scores	5.90	5.8	5.9	3.0	3.4	3.9	3.9	3.9	4.5	**0.70**

*Note***.** The upper triangle refers to correlations between variables, and the bottom triangle refers to the covariances. The bolded values are item variance.

3.4.2 Ascertaining the Number of Profiles

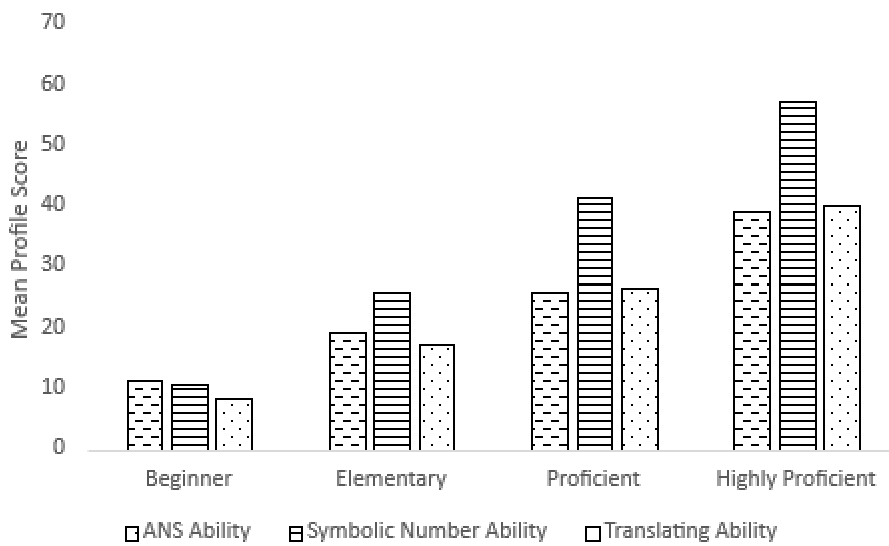
Table 7. AIC, BIC and Entropy of 2-class to 7-class models

Fit Statistic	*2-Class*	*3-Class*	*4-Class*	*5-Class*	*6-Class*	*7-Class*
BIC	12216.39	11459.70	11362.83	11376.44	11444.73	11574.47
ABIC	12143.37	11348.58	11200.92	11151.04	11143.13	11183.98
Entropy	0.827	0.764	0.760	0.813	0.806	0.799

To select the appropriate number of profiles, six exploratory RI-LTAs were fitted to the data. **Table 7** compares the resulting model fit from a two-profile model to a seven-profile model. Examination of fit indices suggest that that a four-profile model or a five-profile model would best fit the data.

Examination of the mean profile scores and profile membership suggest that both the four- profile and five-profile models are theoretically plausible. However, it as found that the results from the five-profile model did not theoretically differ from the results from the four-profile model. As such, the more parsimonious four-profile model was chosen from both a model fit and theoretical interpretability perspective.

Figure 6**.** Four Latent Class Profiles.



3.4.3 What are the Profile Prevalence and Transition Probabilities?

Mean scores for each of the four profiles are illustrated in **Figure 6**. First, examination of mean profile scores revealed that students belonging to profile 1 has the lowest average mean score for symbolic comparison, non-symbolic comparison, and mixed comparison. Thus, this class was the "beginner" class. Second, examination of mean profile scores revealed that students belonging to profile 2 has higher scores than students in profile 1 and lower scores than students in profile 3 or 4 for all three comparison tasks. As such, this class was the "elementary" class. Third, examination of mean profile scores revealed that students belonging to profile 3 high scores than profile 1 or 2 and lower scores than students in profile 4. Thus, this class was the "proficient" class. Finally, examination of mean profile scores revealed that students belonging to profile 4 has the highest average mean score for symbolic comparison, non-symbolic comparison, and mixed comparison. Thus, this class was the "highly proficient" class. Comparisons of mean profile scores for three skills in all four subgroups (e.g., comparing profile 1 ANS scores against the ANS scores of profile 2-4 etc.) are statistically significant (all *p*s < 0.001), this suggest that the same skill in all adject groups significantly differ from one another.

Table 8. Distribution of Profiles and Transition Probabilities

	Profile 1	*Profile 2*	*Profile 3*	*Profile 4*
Profile Probabilities				
Time 1	368.3 (61.4%)	174.4 (29.1%)	57.3 (9.6%)	0 (0%)
Time 2	80.9 (13.5%)	290.8 (48.5%)	200.6 (33.4%)	27.7 (4.6%)
Time 3	12.1 (2.0%)	35.4 (5.9%)	335.2 (55.9%)	217.3 (36.2%)
Transition Probabilities				
Time 1 – Time 2				
Profile 1	.215	.682	.097	.005
Profile 2	.009	.228	.707	.056
Profile 3	.000	.000	.724	.276
Profile 4	.000	.000	.000	1.000
Time 2 – Time 3				
Profile 1	.150	.236	.506	.108
Profile 2	.000	.056	.773	.171
Profile 3	.000	.000	.311	.689
Profile 4	.000	.000	.257	.743

Note. Profile 1 refers to the beginner class, profile 2 refers to the elementary class, profile 3 refers to the proficient class and profile 4 refers to the highly proficient class.

Table 8 reports distribution of profiles and transition probabilities. Baseline distribution of profiles suggest that at T1, 61% of students belonged to the beginner profile and 29% of students belonged to the elementary profile, 10% of students belonged to the proficient profile, and none were in the highly proficient profile. Further, examination of transition probabilities suggest that a vast majority of students transition from a less proficient profile to a more proficient profile over the time period of study. Indeed, by T3, over 80% of students either belonged to the proficient or highly proficient class.

Taken together, given that the four profiles seem to roughly represent a progression from a less proficient state to a more proficient state, and that the transition probabilities between the groups are very high, our results suggest that the four latent profiles may be reflecting levels of development.

3.4.4 Do EQAO Scores Relate with Between-Person Differences and Profile Membership?

EQAO scores were added a distal outcome predicted by the random-intercept and profile membership at T3 to ascertain whether a) between-person differences predict EQAO scores, and b) profile membership at the end of grade 1 predicts EQAO scores at grade 3. Result indicates that the between-person differences as measured by the random-intercept is significantly related with EQAO scores ($B = .481, \beta = .605, SE = .048, p < 0.001$), and that a standard deviation change in the random-intercept score is related with a .481 units change in EQAO score.

The relation between profile membership and EQAO scores are presented in **Table 9**. Interestingly, examination of profile means and standard deviations of EQAO scores suggest that students belonging to the beginner and elementary profiles are more distributed around profile means when compared with the proficient and highly proficient profile. The higher standard deviation values may be a reflection of the low profile membership observed in the beginner and elementary profile at T3. Further, results indicate that students belonging to the beginner profile at grade 1 have significantly lower EQAO score than those students who belonged to the proficient or highly proficient profiles. Similarly, students in the highly proficient profile at grade 1 have significantly higher EQAO score than those in the proficient profile.

Taken together, our results suggest that individual differences in stable, trait-like characteristics (e.g., intelligence, SES, interest) are significantly related with EQAO scores. Further, these results suggest that those children who belonged to the proficient or highly proficient profiles not only achieve higher EQAO scores than their peers who belonged to the beginner class, but they more consistently score around their profile means.

Table 9. Mean EQAO scores for each Profile and Mean Difference Comparisons

	Profile 1	*Profile 2*	*Profile 3*	*Profile 4*
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
EQAO Scores	2.44(1.12)	2.97(1.94)	3.19(1.35)	3.48(1.20)
	EQAO Score Mean Difference (Hedge's g Effect Size)			
Profile 1	-	.534 (.288)	.747(.558)*	1.040(.870)**
Profile 2		-	.213(.152)	.506(.352)*
Profile 3			-	.293(.221)*
Profile 4				-

Note. Profile 1 refers to the beginner class, profile 2 refers to the elementary class, profile 3 refers to the proficient class and profile 4 refers to the highly proficient class. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

3.4.5 Exploratory Analysis: Pace of Development

In the main analyses, we uncovered four latent profiles that exhibit a high degree of transition from a less proficient profile to a more proficient profile. The most common patterns of progression are explored in **Table 10**. In the table, progression refers to students moving from a less proficient profile to a more proficient profile, while regression refers to moving from a more proficient profile to a less proficient profile. Results indicate that a majority of students (58%) progress both from T1 to T2, and T2 to T3, and a significant percentage of students (36%) progress only once from T1 to T3. The prevalence of students who failed to progress from T1 to T3 or regress from T1 to T3 are low, combined they compose of 6% of the population.

3.5 Discussion for Study 2

Through the use of an RI-LTA, the current study ascertained, when between-person influences are controlled for, whether there are subgroups in the population with different patterns of development. Several important findings were uncovered: 1) analysis revealed that there are four distinct subgroups in the population that can best be described as levels of development and that most of the population follow this progression, and 2) that both profile membership and trait-like individual differences predict math achievement scores as measured by EQAO scores.

Table 10. Most Common patterns of progression and Prevalence

Latent Profile Patterns			***Final Profile Estimates***	***Percentage***
T1	***T2***	***T3***		
Double Progression				
Profile 1	Profile 2	Profile 3	193.91	32.23%
Profile 2	Profile 3	Profile 4	84.88	14.14%
Profile 1	Profile 2	Profile 4	42.97	7.16%
Others Double Progression Patterns			24.71	4.12%
Total			346.47	57.74%
Stationary T1 – T2				
Profile 2	Profile 2	Profile 3	42.93	7.15%
Profile 1	Profile 1	Profile 3	35.89	5.98%
Profile 1	Profile 1	Profile 2	31.52	5.25%
Other Stationary T1 – T2 Patterns			28.02	4.67%
Total				22.87%
Stationary T2 – T3				
Profile 2	Profile 3	Profile 3	40.27	6.71%
Profile 1	Profile 2	Profile 2	19.86	3.31%
Profile 1	Profile 3	Profile 3	16.31	2.71%
Other Stationary T2 – T3 Patterns			5.73	0.95%
Total				13.68%
No Progression				
Profile 3	Profile 3	Profile 3	12.92	2.15%
Profile 1	Profile 1	Profile 1	11.89	1.98%
Profile 2	Profile 2	Profile 2	2.24	0.37%
Total			27.05	4.51%
Regression				
Profile 3	Profile 4	Profile 3	4.06	0.68%
Profile 2	Profile 4	Profile 3	2.53	0.42%
Profile 2	Profile 1	Profile 3	0.76	0.12%
Other Regression Patterns			1.26	0.21%
Total			8.61	1.44%

Note. Profile 1 refers to the beginner class, profile 2 refers to the elementary class, profile 3 refers to the proficient class and profile 4 refers to the highly proficient class.

3.6.1 Levels of Development

In the main analyses, examination of children's ANS ability, symbolic number ability, and translating ability revealed that there are four distinct subgroups in the population that can best be described as "beginner," "elementary," "proficient," and "highly proficient." Examination of the transition probabilities suggest that from senior kindergarten to grade 1, most students progress from a lower proficiency profile to a higher proficiency profile. Taken together, our results suggest that these four profiles best reflect levels of development that children undergo. As our dataset did not include reaction times of participants, we could not fully replicate the analyses done by Chew et al. (2016). However, at least with respect to performance, these results stand in contrast to previous findings suggesting that there are subgroups in the population with different developmental profiles (Chew et al., 2016).

However, while our results support the notion that there is homogeneity in the levels of development children undergo, our results also indicate that there is heterogeneity in the pace of development. More specifically, while we observed a large portion of students (56%) consistently progress from a lower performance profile to a higher performance profile over T1 to T3, a significant portion of students (36%) only progress to a higher performance profile once over the three timepoints or had seen no progression at all (5%). In other words, our results have shown that approximately 40% of the population show some degree of inconsistent progress in skills from senior kindergarten to grade 1.

This heterogeneity in the pace of development suggests that estimates of longitudinal associations between variables (e.g., associations between earlier ANS ability and later math achievement) could be highly varied depending on the sample size. If a smaller sample size is chosen, it is easier for the chosen group of participants to have higher proportions of students with inconsistent progression or no progression by chance, leading to less accurate parameter estimations.

In sum, our results suggest that while there is homogeneity in the developmental trajectories that children undergo, there is some degree of heterogeneity in terms of the pace children progress through the levels of development.

3.6.2 Levels of Development and Developmental Dyscalculia

Interestingly, these results may also have implications on research examining students with developmental dyscalculia (DD), a persistent and specific disorder that is characterized by impaired mathematical development and learning despite possessing normal intelligence (Butterworth, 2008). First, given that the best fitting model did not include a low-performing profile that is isolated (i.e., having low transition probabilities into and out of other profiles), our results suggest that the performance profiles (i.e., ANS ability score, symbolic number ability score, and translating ability score) of children with DD and typically developing children may be fairly similar at T1 (the beginning of senior kindergarten). This is in line with the hypothesis that DD arises from an diminished ability to represent numerical magnitude causing an impairment of numerical learning (Butterworth, 1999, 2008; Wilson & Dehaene, 2007). While typically developing children would progress to a more proficient profile over time, children with DD would continue to stay in the beginner profile.

Second, our results also suggest that there may be a risk of misdiagnosing children as having DD, if performance in only two timepoints are used as reference for diagnosis. Indeed, in our sample, only 2% of children stayed in the beginner profile from T1 to T3. In contrast, approximately 12% of children stayed in the beginner profile for T1 and T2 but moved to a more proficient profile in T3. With only information from T1 and T2, the probability of selecting a child who will stay in the beginner profile in T3 is only 17%.

Taken together, our results suggest that there are no distinct subgroups in the population that exhibits different developmental profiles. Instead, we find that most children generally follow the levels of development at different paces, and that children with DD do not progress beyond the beginner profile.

3.6.3 Predicting Math Achievement Two Years Later

A second finding of the current study is that both the between-person variability (as measured by the random-intercept) and profile membership at T3 significantly predicts students' EQAO scores at Grade 3. When relating between-person variability in time-invariant trait-like characteristics (e.g., some combination of general math ability, intelligence, interest in mathematics, SES, etc.) to EQAO scores, results indicate that a one standard deviation difference in the random-intercept corresponds to .481 units change in EQAO scores.

While results from mean comparisons were somewhat inconclusive (see **Table 9**), results seem to suggest that there are significant mean differences in EQAO scores between the lower performers (beginner and elementary profiles) and the high performers (proficient and highly proficient profiles). Further, there was also statistically significant differences in mean EQAO scores between the proficient and highly proficient profiles.

One way of comparing the effect sizes of the within-person effect and between-person effect is to compare the profile mean difference in EQAO scores versus the change in EQAO score per 1 standard deviation change in the random intercept. On average, mean differences between adjacent profiles (beginner vs. elementary, elementary vs. proficient, and proficient vs. highly proficient) was 0.347. Mean difference between profiles separated by one profile (beginner vs. proficient and elementary vs. highly proficient) was 0.627. Finally, mean difference between the lowest and highest profile was 1.040. This is compared with .481 unit change in EQAO score per 1 standard deviation change in the random intercept.

On average, a movement from a lower proficiency profile to an adjacent, more proficient profile is associated with change in EQAO scores that is somewhat smaller than a 1 standard deviation change in the random intercept. While, on average, a movement from the beginner to proficient profile or elementary to highly proficient profile is associated with change in EQAO scores that is comparable to a 1 standard deviation change in the random intercept. Taken together, our results suggest that both between-person differences, and within-person development come together to influence students' future math achievement. Further, comparison of within-person and between-person effect suggests that both account for similar amounts of variability in EQAO scores.

3.7 Conclusions

Using an RI-LTA analysis, the current study addressed the issue of whether existing accounts of numerical development may be incomplete due to extant studies demonstrating subgroups in the population. Results suggests that a vast majority of students seem to follow the same developmental trajectory. As such, existing accounts of numerical development seem to adequately describe children's numerical development.

3.8 References

Bartelet, D., Ansari, D., Vaessen, A., & Blomert, L. (2014). Cognitive subtypes of mathematics learning difficulties in primary education. Research in Developmental Disabilities, 35(3), 657–670.

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

Butterworth, B. (1999). The mathematical brain. Macmillan.

Butterworth, B. (2008). Developmental dyscalculia. Child Neuropsychology: Concepts, Theory, and Practice, 357–374.

Carey, S. (2001). Cognitive foundations of arithmetic: Evolution and ontogenisis. Mind & Language, 16(1), 37–55.

Carey, S. (2004). Bootstrapping & the origin of concepts. Daedalus, 133(1), 59–68. https://doi.org/10.1162/001152604772746701

Carey, S. (2009). The Origin of Concepts. Oxford University Press.

Chew, C. S., Forte, J. D., & Reeve, R. A. (2016). Cognitive factors affecting children's nonsymbolic and symbolic magnitude judgment abilities: A latent profile analysis. Journal of Experimental Child Psychology, 152, 173–191.

Chew, C. S., Forte, J. D., & Reeve, R. A. (2019). Implications of Change/Stability Patterns in Children's Non-symbolic and Symbolic Magnitude Judgment Abilities Over One Year: A Latent Transition Analysis. Frontiers in Psychology, 10, 441.

Clayton, S., & Gilmore, C. K. (2015). Inhibition in dot comparison tasks. Zdm, 47(5), 759–770.

Collins, L. M., & Lanza, S. T. (2009). Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences (Vol. 718). John Wiley & Sons.

Curfman, G. D. (1993). Is exercise beneficial—Or hazardous—To your heart? Mass Medical Soc.

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. Annual Review of Psychology, 62, 583–619.

de Souza Salvador, L., Moura, R., Wood, G., & Haase, V. G. (2019). Cognitive heterogeneity of math difficulties: A bottom-up classification approach. Journal of Numerical Cognition, 5(1), 55–85.

Dehaene, S. (1999). The Number Sense: How the Mind Creates Mathematics. Oxford University Press.

Dehaene, S. (2005). Evolution of human cortical circuits for reading and arithmetic: The "neuronal recycling" hypothesis. From Monkey Brain to Human Brain, 133–157.

Dehaene, S. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In Sensorimotor foundations of higher cognition (pp. 527–574).

Diez-Roux, A. V. (1998). Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. American Journal of Public Health, 88(2), 216–222.

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

Ferreira, F. de O., Wood, G., Pinheiro-Chagas, P., Lonnemann, J., Krinzinger, H., Willmes, K., & Haase, V. G. (2012). Explaining school mathematics performance from symbolic and nonsymbolic magnitude processing: Similarities and differences between typical and low-achieving children. Psychology & Neuroscience, 5, 037–046.

Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. Structural Equation Modeling: A Multidisciplinary Journal, 4(2), 87–107. https://doi.org/10.1080/10705519709540063

Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low‐income homes: Contributions of inhibitory control. Developmental Science, 16(1), 136–148.

Gilmore, C. K., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., Simms, V., & Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. PloS One, 8(6), e67374.

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.

Hickendorff, M., Edelsbrunner, P. A., McMullen, J., Schneider, M., & Trezise, K. (2018). Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis. Learning and Individual Differences, 66, 4–15.

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. Psychological Methods, 11(1), 36.

Hoffman, L., & Stawski, R. S. (2009). Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. Research in Human Development, 6(2–3), 97–120.

Kißler, C., Schwenk, C., & Kuhn, J.-T. (2021). Two dyscalculia subtypes with similar, low comorbidity profiles: A mixture model analysis. Frontiers in Psychology, 12, 1828.

Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. (2013). Early numerical development and the role of non-symbolic and symbolic skills. Learning and Instruction, 25, 95–103.

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool Acuity of the Approximate Number System Correlates with School Math Ability. Developmental Science, 14(6), 1292–1300. https://doi.org/10.1111/j.1467-7687.2011.01080.x

Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Is approximate number precision a stable predictor of math ability? Learning and Individual Differences, 25, 126–133.

Lyons, I. M., Bugden, S., Zheng, S., De Jesus, S., & Ansari, D. (2018). Symbolic number skills predict growth in nonsymbolic number skills in kindergarteners. Developmental Psychology, 54(3), 440. https://doi.org/10.1037/dev0000445

Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. PLoS One, 6(9), e23749.

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). Quantitative development in infancy and early childhood. Oxford University Press.

Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. Journal of Experimental Child Psychology, 103(4), 490–502. https://doi.org/10.1016/j.jecp.2009.02.003

Muthén, B., & Asparouhov, T. (2020). Latent Transition Analysis With Random Intercepts (RI-LTA).

Muthén, L. K., & Muthén, B. O. (2010). Mplus: Statistical analysis with latent variables: User's guide. Muthén & Muthén Los Angeles.

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

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. https://doi.org/10.1371/journal.pone.0067918

Paglin, M., & Rufolo, A. M. (1990). Heterogeneous human capital, occupational choice, and male-female earnings differences. Journal of Labor Economics, 8(1, Part 1), 123–144. https://doi.org/10.1086/298239

Parsons, S., & Bynner, J. (2005). Does numeracy matter more?

Peel, D., & McLachlan, G. J. (2000). Robust mixture modelling using the t distribution. Statistics and Computing, 10(4), 339–348.

Piazza, M. (2011). Neurocognitive start-up tools for symbolic number representations. Space, Time and Number in the Brain, 267–285.

Price, G. R., & Fuchs, L. S. (2016). The Mediating Relation between Symbolic and Nonsymbolic Foundations of Math Competence. PLOS ONE, 11(2), e0148981. https://doi.org/10.1371/journal.pone.0148981

Rivera-Batiz, F. L. (1992). Quantitative literacy and the likelihood of employment among young adults in the United States. Journal of Human Resources, 313–328. https://doi.org/10.2307/145737

Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. British Journal of Developmental Psychology, 30(2), 344–357.

Sasanguie, D., Van den Bussche, E., & Reynvoet, B. (2012). Predictors for mathematics achievement? Evidence from a longitudinal study. Mind, Brain, and Education, 6(3), 119–128.

Scholl, B. J. (2001). Objects and attention: The state of the art. Cognition, 80(1–2), 1–46.

Simon, T. J. (1997). Reconceptualizing the origins of number knowledge: A “non-numerical” account. Cognitive Development, 12(3), 349–372.

Szucs, D., Nobes, A., Devine, A., Gabriel, F. C., & Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. Frontiers in Psychology, 4, 444.

Wilson, A. J., & Dehaene, S. (2007). Number sense and developmental dyscalculia.

Chapter 4

4.1 Introduction

Math anxiety is the "feeling of tension, apprehension or even dread, that interferes with the ordinary manipulation of numbers and the solving of mathematical problems" (Ashcraft & Faust, 1994, p.98). Consistent and robust associations have been demonstrated between math anxiety and math achievement, indicating that people with higher feelings of fear and anxiety towards math tend to have lower math achievement (Ashcraft & Krause, 2007; Barroso et al., 2020; Hembree, 1990; Ma, 1999). Results from a recent meta-analysis by Barroso et al. (2020) estimate the correlation to be $r = 0.28$, a small-to-moderate effect size, comparable to the effects of other important correlates of math achievement, including magnitude processing skills ($r = 0.24$; Chen & Li, 2014) and socioeconomic status ($r = 0.35$; Sirin, 2005).

High prevalence rates of math anxiety have been reported across countries (Lee, 2009b) and age groups (Ma, 1999; Vukovic et al., 2013), and the consequences of math anxiety are far reaching. People with heightened levels of math anxiety often experience a lifelong tendency to avoid math, math-related situations, career paths that require math, and most notably, courses and degrees in STEM (Ashcraft & Ridley, 2005; Chipman et al., 1992; Meece et al., 1990). In short, there is evidence that math anxiety negatively impacts math performance and can influence how one experiences and interacts with the world.

4.1.1 Math Anxiety's Effects on Math Achievement

An in-depth examination of the association between math anxiety and math achievement suggests that math anxiety is detrimental to performance in many school-related math skills, including problem-solving (Hembree, 1990), simple arithmetic (Ashcraft & Faust, 1994) and basic number processing (Maloney, Ansari, & Fugelsang, 2011). The most influential account of the mechanism behind these associations posits that math anxiety interferes with math performance by compromising cognitive resources that are key for success in math. Specifically, worries and intrusive thoughts evoked by math anxiety disrupt and compete for cognitive resources, such as working memory, that are vital for mathematics problem solving (Ashcraft et al., 1992).

Both behavioral (Ashcraft & Faust, 1994b; Faust, 1996; Ramirez et al., 2013, 2016; Vukovic et al., 2013) and neuroimaging (Lyons & Beilock, 2011; Pletzer et al., 2015; Sarkar et al., 2014) evidence has been forwarded in support of this interference account.

4.1.2 The Predictors of Math Anxiety

Though the characteristics of math anxiety and its consequences has been relatively well described, the predictors of math anxiety are multifaceted and less well understood. Prior studies have identified a large repertoire of individual and environmental factors associated with math anxiety. Individual factors include, genetics (Wang et al., 2014), working memory capacity (Ramirez et al., 2013) attentional bias (Rubinsten et al., 2015) and affective or physiological responses (Lyons & Beilock, 2012). Environmental factors include parental support and expectation (Vukovic et al., 2013), parental attitudes towards mathematics (Soni & Kumari, 2017b) math activities at home (Berkowitz et al., 2015), classroom atmosphere (Fast et al., 2010b), teacher's own math anxiety (Beilock et al., 2010b; Schaeffer et al., 2020), teacher self-efficacy (Swars et al., 2006), teacher expectations of students (Mizala et al., 2015b) and cultural background (Foley et al., 2017).

4.1.3 Limitations of the Extant Literature

The abovementioned studies have markedly increased understanding of the individual and environmental variables that contribute to math anxiety and of the mechanisms through which math anxiety affects math achievement. However, these studies are nevertheless limited in three important and interconnected aspects that have stifled understanding of math anxiety.

4.1.4 Information Silos

First, the extant literature examining math anxiety have largely operated in separate information silos with little crosstalk; this consequently has left several important questions unanswered. Specifically, while multiple individual and environmental factors have been found to be related to math anxiety, most studies that examine these individual and environmental factors have done so in relative isolation of other proposed predictors of math anxiety. For example, the basic numerical magnitude processing predictors of math anxiety have not been studied in tandem with environmental factors, such as teachers' math anxiety. As such, little is known regarding whether any of the known environmental and individual predictors would uniquely predict math anxiety after controlling for other proposed predictors of math anxiety. Moreover, studies that examine, on the one hand, the predictors of math anxiety and, on the other, the relationship between math anxiety and math achievement, have also been conducted in isolation of one another. As such, little is known regarding whether the identified individual and environmental predictors of math anxiety may serve as unobserved confounders that may inflate the relation between math anxiety and math achievement.

4.1.5 Multilevel Nature of Math Anxiety

Second, perhaps as a consequence of information silos in the extant literature, the multilevel nature of math anxiety has been scarcely explored. Particularly, hints of the multilevel nature of math anxiety can be found from studies that showed multiple environmental predictors of math anxiety (e.g., classroom atmosphere (Fast et al., 2010b) and teacher's own math anxiety (Beilock et al., 2010b; Schaeffer et al., 2020)) exist at a higher, clustered level (i.e., these environmental factors affect a group of individuals who shared the same education environment rather than specific individuals). This suggests there may be systematic variations of math anxiety at the education environment level, and student membership in any particular education environment would be predictive of student math anxiety to some degree (please see methods for reasons why we elected to use the broader term education environment to refer to influences from the school or classroom).

Importantly, when considering math anxiety as a predictor, the variability of math anxiety at the individual level and variability of math anxiety at the education environment level may have independent and different effects on student math achievement. To see why this might be the case, it is useful to envision a scenario where a child with some degree of math anxiety is put into an education environment with low average math anxiety, and an alternative scenario where the same child is put into an education environment with high average math anxiety. Irrespective of the cause of a particular environment's average level of math anxiety (i.e., classroom atmosphere (Fast et al., 2010b) and teacher's own math anxiety (Beilock et al., 2010b; Schaeffer et al., 2020)), it is possible that the environment average level of math anxiety could serve to predict the child's math achievement over and above what could be predicted by the child's own level of math anxiety.

Put more formally, the effect that the education environment-average math anxiety has on individual math achievement is a contextual effect. A contextual effect is said to occur if an aggregation (i.e., mean) of an individual-level variable at a higher level (i.e., aggregating individual math anxiety by calculating classroom average math anxiety) makes an independent contribution to explaining the outcome variable over and above the contribution of the same variable at the lower level (i.e., individual math anxiety). In this case where math anxiety is used to predict math achievement, a contextual effect of math anxiety at the education environment level is the additional variance of math achievement explained by education environment-average math anxiety over and above the variance explained by individual math anxiety.

Decomposing the total effect of math anxiety on math achievement into the individual effect and contextual effect allows for a more accurate account of how math anxiety affects math performance. Indeed, the exploration of the potential of contextual effects is especially important, as it is possible for contextual effects to significantly differ in both size and direction from the individual effect. For instance, research on the big-fish-little-pond effect (see Marsh et al., 2008 for a review) has revealed that while the relationship between achievement and self-efficacy may be positive at the individual-level (i.e., the better my grades, the more confident I am in my abilities), the relationship is reversed at the classroom-level (i.e., the better my peer's average grades, the less confident I am in my abilities).

To date, few studies (e.g., Beilock et al., 2010; Schaeffer et al., 2020) have accounted for the multilevel nature of math anxiety, and to our knowledge, no studies thus far have examined the possibility of a contextual effect of math anxiety at the education environment level. Failing to account for the multilevel nature of math anxiety is problematic, statistically, theoretically, and practically. Statistically, members in the same group are more likely to be similar than members from a different group. Analyses that ignore this grouping structure violates the assumption of independence, thereby increasing the risk of yielding misleading parameter estimates, standard error estimates, fit indices, and the risk of Type I error (Julian, 2001; Kreft & Yoon, 1994). Theoretically, as it is possible for individual and contextual effects to differ in both strength and direction, conclusions drawn using analyses that ignore the nested nature of the data is incomplete and often misleading (Diez-Roux, 1998). Practically, as no study thus far has examined the size and direction of the contextual effect at the education environment, the utility of interventions for the alleviation or prevention of the adverse effects of math anxiety at different levels (e.g., individual intervention vs. classroom intervention) is currently unknown.

4.1.6 An International Perspective

Third and finally, the overrepresentation of participants from western, educated, industrialized, rich and democratic (WEIRD) societies is a prevalent and important problem afflicting research in psychology (Henrich et al., 2010). For instance, in one study, it was found that over 96% of subjects in papers published in top psychological journals belong to WEIRD societies (Arnett, 2008). This problem similarly applies to the extant literature on math anxiety, and much is still unknown regarding math anxiety in other societies.

When put into a multilevel modelling context, there may be between-country differences in in the magnitude of effect that math anxiety has on math achievement, both as an individual effect (i.e., the amount of increase or decrease in math achievement per unit increase in individual math anxiety) and as a contextual effect (i.e., the amount of increase or decrease in math achievement per unit increase in education environment-average math anxiety). Further, similar to the contextual effect of math anxiety at the education environment level, it is also possible to examine whether there is a contextual effect of math anxiety at the country level. Going back to the hypothetical situation outlined above, a contextual effect at the country level would entail a scenario where the same child is put into a low-average math anxiety country and an alternative scenario where the same child is put into a high-average math anxiety country. If the country's average level of math anxiety is predictive of math achievement beyond what could be accounted for by the child's individual math anxiety, there would be a contextual effect of math anxiety at the country level.

Thus far, only a few studies in the literature have systematically explored the possibility of between-country differences in math anxiety (Foley et al., 2017; Lee, 2009b). Critically, no studies have explored between-country differences in the individual and contextual effects of math anxiety.

4.1.7 Study 3

To fill in the research gap identified above, the current study utilized a multilevel structural equation modelling (ML-SEM) approach in order to model the relations between the math anxiety, math achievement, and their predictors. The current study draws on data from three large-scale international studies of student achievement: Trends in International Mathematics and Science Study 2015 Grade 4 and Grade 8 (TIMSS Grade 4 and Grade 8; Martin, Mullis, & Hooper, 2016) and the Programme for International Student Assessment 2012 Grade 8 (PISA; OECD, 2014). This sample is larger and more diverse than any other math anxiety study to date. The current study extends existing research in three important ways:

1. Given the dearth of research that examine the relation between math anxiety and math achievement under a multilevel modelling framework, we investigate whether the aggregation of math anxiety at the education environment (i.e., education environment-average math anxiety) makes an independent contribution to explaining math achievement over and above the contribution of individual math anxiety. Further, few extant studies have systematically examined the potential of between-country differences in the individual and contextual effects. The exploration of possible between-country differences in the individual and contextual effects in the current study is important, as it would highlight the degree to which research findings can be applied in different cultural contexts.

2. Extant research on the predictors of math anxiety have been done in relative isolation (e.g., cognitive predictors of math anxiety (e.g., Maloney et al., 2012) have typically been studied separately from environmental predictors of math anxiety, such as the home environment (e.g., Soni & Kumari, 2017)). As the TIMSS Grade 4 database include a rich set of potentially relevant predictors of math anxiety (see methods), we investigated the relative strength of influence of these variables. In doing so, we consolidate current research findings and provide a more coherent picture of how different variables may predict math anxiety.

3. Since many factors have been found to predict math anxiety, it is important to examine how these predictors may affect the relationship between math anxiety and math achievement. Including these predictors alongside the estimation of math anxiety's effect on math achievement

would control for potential shared variance explained between the predictors and math anxiety. We therefore explore whether any of these variables may serve as an unobserved confounder that inflate the relation between math anxiety and math achievement.

4.2 Methods

4.2.1 Data Sources - TIMSS Grade 4, Grade 8, and PISA

Study 3 draws upon the TIMSS 2015 Grade 4 and Grade 8 databases and PISA 2012 database. While a more recent PISA database is available, the most contemporary database does not include the measurements of math anxiety. The sample design for the TIMSS and the PISA databases is a stratified two-stage random sample design (Martin et al., 2016; OECD, 2014). Both studies draw a sample of schools from participating countries as a first stage. As a second stage, one intact class of students is selected from each of the sampled schools in the case of the TIMSS (M. O. Martin et al., 2016), and a random selection of eligible students that are not necessary from the same class are drawn in the case of the PISA (OECD, 2014a). Consequently, the second level of nesting can be most aptly described as the classroom-level for the TIMSS and school-level for the PISA. For current purposes, we will refer to this second-level as the education environment level. The differences in data collection method cases the statistical interpretation of the results is somewhat different between the databases (see **Appendix 5** for more details and other statistical considerations).

4.2.2 Participants

Combining all three databases, the total sample size exceeded 1 million participants (*N*=1,175,515). The final sample for the TIMSS Grade 4 database included 404,688 students (196,412 females; mean age=10.11; SD=.61) from 21,600 classrooms (mean number of students per classroom = 18.74), across 54 countries (mean number of classrooms per country=7,494.22). The final sample for the TIMSS Grade 8 included 290,653 students (144,277 females; mean age=14.23; SD = .79) from 14,426 classrooms (mean number of students per classroom=20.15), across 46 countries (mean number of classrooms per country=6,318.54). The final sample for the PISA database included 480,174 students (242,375 females; mean age=15.78; SD=.29), from 18,139 schools (mean number of students per school=26.47), across 65 countries (mean number of schools per country=7387.29).

4.2.3 Math Achievement

Across all three databases, math achievement was measured through a comprehensive test, targeting a wide variety of mathematical skills and concepts. The TIMSS Grade 4 assessment measures topics related to number, geometry shapes and measure, and data display. The TIMSS Grade 8 assessment measures topics related to number, algebra, geometry, data and chance (M. O. Martin et al., 2016). The PISA measures topics related to change and relationships, quantity, space and shape and uncertainty and data (OECD, 2014a). Studies that have compared math achievement assessed by the TIMSS and PISA have found that questions in the TIMSS tend to be more theoretically oriented and questions in PISA tend to more application oriented (Hutchison & Schagen, 2007).

In both TIMSS and the PISA, math achievement score of students was measured using a rotating booklet design whereby students complete only a subset of all assessment items (M. O. Martin et al., 2016; OECD, 2014a). As such, some degree of measurement error is introduced by this method of assessment (Von Davier et al., 2009). As a reflection of this uncertainty introduced by the rotating booklet design, five plausible values are provided by the TIMSS and PISA databases for each individual as a representation of proficiency. Correct analyses of plausible values require separate identical analyses for each plausible value with results integrated using principles from multiple imputation.

4.2.4 Math Anxiety

As part of all three achievement tests, students were asked to complete a series of self-report measures, including questions about math anxiety. On both the Grade 4 and Grade 8 TIMSS assessment, students were asked to indicate on a 4-point Likert scale (1=agree a lot…4=disagree a lot) the extent to which they agreed with the statement: "*Mathematics makes me nervous*." For the PISA, math anxiety was operationalized an observed variable calculated as the mean score of 5 individual test items. Students were asked to indicate on a 4-point Likert scale the extent to which they agreed with each of the following statements (1=strongly agree…4=strongly disagree): "*I* *often worry that it will be difficult for me in mathematics classes*," "*I get very tense when I have to do mathematics homework*," "*I get very nervous doing mathematics problems*," "*I feel helpless when doing a mathematics problem*," and *"I worry that I will get poor grades in maths."* All items were reverse coded where appropriate.

4.2.5 Predictors of Math Anxiety and Math Achievement

Predictors of math anxiety were extracted from the TIMSS Grade 4 database. It was not possible to investigate the predictors of math anxiety in the other two databases due to unavailable data. Specifically, TIMSS Grade 8 did not collect any data regarding the students' home and parents, and the PISA did not collect any data regarding the teacher due to the study design.

The TIMSS Grade 4 provide a rich set of potentially relevant variables that may affect math anxiety and math achievement, study 3 takes full advantage of this by including these variables as predictors of math anxiety and math achievement across three levels of analysis: *individual level*, *education environment level*, and *country level*.

At the *individual level* (L1), variables were sub-divided into three categories: (1) student specific factors, (2) past and present extracurricular mathematics training, and (3) parent and home factors. Student specific factors included student self-report items related to student gender, student attitudes towards mathematics, and student attitudes towards the school. Past and present extracurricular mathematics training included parent's self-report items related to years of preschool education, home mathematics activities during preschool, and current extracurricular tutoring/lessons. Parent and home factors included parent's self-report items related to parental involvement in mathematics homework, parental attitudes towards mathematics and science, parents' highest education level, parents' occupation and home socioeconomic status.

At the *classroom level* (L2), variables were sub-divided into two categories: (1) Classroom factors and (2) teacher specific factors. Classroom factors include weekly time spent on mathematics, frequency of mix-ability grouping, frequency of same-ability groups and math homework frequency. Teacher specific factors include teacher gender, teacher satisfaction with work, teacher confidence in teaching mathematics, teacher years of teaching experience, and teacher major.

Finally, at the *country level* (L3), we included socioeconomic development and cultural dimensions from Hofstede et al. (2010). Socioeconomic development was represented with the U.N. Human Development Index (HDI; United Nations Development Program, 2015). Cultural differences between countries were represented by five dimensions (individualism-collectivism, power distance, uncertainty avoidance, masculinity, and long-term orientation) as proposed by Hofstede et al. (2010). See **Appendix 6** for a description of the variables. All values were obtained from Hofstede et al. (2010). The HDI is an index that account for multiple facets of human development a specific country, this includes gross domestic product per capita, life expectancy, adult literacy rate and school enrollment ratio. For study 3, we used the values from 2015, the year of data collection for the TIMSS Grade 4.

All predictors were reverse coded and operationalized as latent variables where appropriate. For the substantive basis for the included variables, and details of the item wordings of main variables and covariates please see **Appendix 6** and **Appendix 7** in the supplemental material.

4.3 Statistical Analyses

Missing data were handled using multiple imputation (Graham & Hofer, 2000; Schafer, 1999). Analyses were carried out using Mplus 8.3 with the maximum likelihood estimator with robust standard errors (L. K. Muthén & Muthén, 2010). All continuous variables were standardized (mean=0, SD=1) prior to estimation to remove non-essential multicollinearity (Marsh et al., 2012). For all models below, math anxiety as a predictor was grand-mean centered, and a manifest aggregation approach was used to estimate the contextual effect (Marsh et al., 2009). This implies that the higher-level regression coefficients are a direct estimation of the contextual effect that controls for lower-level variations (Brincks et al., 2017; Enders & Tofighi, 2007). Effect sizes for L1-L3 effects were calculated according to Marsh et al. (2009):

$$\Delta = 2 * \beta * \sigma_{pred} / \sigma_y$$

Where β is the unstandardized regression coefficient, σ_{pred} is the standard deviation of the predictor variable, and σ_y is the standard deviation of the outcome variable. This effect size metric is comparable with Cohen's *d* (Cohen, 2013).

Where applicable, goodness of fit of the models were assessed with χ^2 test statistic, the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI) and the root mean square error of approximation (RMSEA). Typical cut-off scores for excellent and adequate fit are CFI and TLI > .95 and > .90, and RMSEA <.06 and <.08 (Hu & Bentler, 1999). More details regarding weighting, confirmatory factor analysis, and assessment of model fit are given in the supplemental materials.

4.3.1 Is there a Contextual Effect of Math Anxiety at the Education Environment Level?

We first sought to establish a baseline model by employing a total group analysis using a two-level model for each database to account for the nesting of students in the immediate education environment. This baseline model will yield estimates of the L1 individual effect and ascertain whether there is evidence for a L2 contextual effect. Adopting the notation used by Raudenbush & Bryk (2002), we have the following model:

$$\text{L1: } (Y_{i,j,k} - \beta_{0,0,k}) = \pi_{0,j,k} + \pi_{1,j,k}\left(X_{i,j,k} - \bar{X}_{\cdot,\cdot,k}\right) + e_{i,j,k}$$

$$\text{L2: } (\pi_{0,j,k} - \beta_{0,0,k}) = \beta_{0,1,k}\left(X_{\cdot,j,k} - \bar{X}_{\cdot,\cdot,k}\right) + r_{0,j,k}$$

Where the variable $Y_{i,j,k}$ is the math achievement for person *i* in education environment *j* in country *k*. The predictors individual math anxiety ($X_{i,j,k}$) and education environment-average math anxiety ($X_{\cdot,j,k}$) are centered with respect to the country means ($\bar{X}_{\cdot,\cdot,k}$). To account for the fact that education environments are nested into countries, country membership was treated as a stratification variable, as such, standard errors and test statistics were corrected for the nesting of education environment within country. It is noted that since math achievement was centered around country means ($\beta_{0,0,k}$), we have removed the between-country variation in math achievement that would be have been otherwise attributed to L2 (Van den Noortgate et al., 2005).

4.3.2 Are there Between-Country Differences in the Contextual Effect at the Education Environment level?

Next, to ascertain whether the baseline model generalizes to all countries, a multigroup two-level model for each database was estimated. Specifically, we modelled the same L1 individual effect and L2 contextual effect as the previous analysis. Country membership was treated instead as a fixed-effect grouping variable. To test for between-country differences in the magnitude of the individual and contextual effects, two multigroup two-level models were compared for each database: 1) an unconstrained model in which the structural parameter at L1 and L2 are allowed to vary across countries and 2) a constrained model in which the structural parameters at L1 and L2 are held constant across countries. A significant reduction in model fit when structural parameters are held constant across countries will suggest that there are significant between-country differences in the magnitude of the individual and contextual effects.

4.3.3 What Individual and Environmental Factors predict Math Anxiety?

The following analyses was performed only on TIMSS Grade 4 due to lack of available data in the other two databases. To examine whether the same individual and environmental factors may predict math anxiety, we estimated a three-level model with math anxiety as the outcome variable. Math anxiety is grand-mean centered. L1, L2 and L3 predictors were also added:

$$\text{L1: } Y_{i,j,k} = \pi_{0,j,k} + \pi_{1,j,k}(\text{STU GENDER})_{i,j,k} + \cdots + e_{i,j,k}$$

$$\text{L2: } \pi_{0,j,k} = \beta_{0,0,k} + \beta_{0,1,k}(\text{TEA GENDER})_{.,j,k} + \cdots + r_{0,j,k}$$

$$\text{L3: } \beta_{0,0,k} = \gamma_{0,0,0} + \gamma_{0,0,1}(\text{HDI})_{.,.,k} + \cdots + \mu_{0,0,k}$$

4.3.4 Are the Individual and Contextual Effects of Math Anxiety Robust to Other Predictors of Math Achievement?

The following analyses was performed only on TIMSS Grade 4 due to lack of available data in the other two databases. A possible area of between-country differences in the relations between math anxiety and math achievement is between-country differences in the magnitude of the L1 individual effect and L2 contextual effect. To examine this, we estimated a three-level model with random slopes above and included L1, L2 and L3 predictors into the model:

$$\text{L1: } Y_{i,j,k} = \pi_{0,j,k} + \pi_{1,j,k}\left(X_{i,j,k} - \bar{X}_{.,.,.}\right) + \pi_{2,j,k}(\text{STU GENDER})_{i,j,k} + \cdots + e_{i,j,k}$$

$$\text{L2: } \pi_{0,j,k} = \beta_{0,0,k} + \beta_{0,1,k}\left(X_{.,j,k} - \bar{X}_{.,.,.}\right) + \beta_{0,2,k}(\text{TEA GENDER})_{.,j,k} + \cdots + r_{0,j,k}$$

$$\pi_{1,j,k} = \beta_{1,0,k} + \beta_{1,2,k}(\text{TEA GENDER})_{.,j,k} + \cdots + r_{1,j,k}$$

$$\text{L3: } \beta_{0,0,k} = \gamma_{0,0,0} + \gamma_{0,0,1}(X_{.,.,k} - \bar{X}_{.,.,.}) + \gamma_{0,0,2}(\text{HDI})_{.,.,k} + \cdots + \mu_{0,0,k}$$

$$\beta_{0,1,k} = \gamma_{0,1,0} + \gamma_{0,1,1}(\text{HDI})_{.,.,k} + \cdots + u_{0,1,k}$$

$$\beta_{1,0,k} = \gamma_{1,0,0} + \gamma_{1,0,1}(\text{HDI})_{.,.,k} + \cdots + u_{1,0,k}$$

Math anxiety and individual and environmental predictors from the previous analysis are added as predictors of math achievement. In doing so, we control for the influences of these predictors and examine whether the variance explained by math anxiety overlap with that accounted for by other predictors. The individual effect ($\pi_{1,j,k}$) and contextual effect ($\beta_{0,1,k}$) of math anxiety are modelled as random slopes, and we regressed the L2 and L3 random slopes onto the L2 and L3 predictors. This will allow us to examine whether these individual and environmental factors may account for the between-environment and between-country differences in the magnitude of effect of the individual and contextual effects.

4.4 Results

Results most pertinent to addressing the three research questions are presented below. See **Appendix 9** for descriptions and results of additional analyses.

4.4.1 Is there a Contextual Effect of Math Anxiety at the Education Environment Level?

To establish a baseline model, a total sample model was computed while ignoring country membership. In the total sample of TIMSS Grade 4, L1 individual math anxiety was negatively related to math achievement ($\pi_{1,j,k} = -.220, SE = .004, \Delta = .524$), and the contextual effect of L2 math anxiety was also negative ($\beta_{0,0,k} = -.186, SE = .013, \Delta = .445$). In the total sample of TIMSS Grade 8, L1 individual math anxiety was negatively related to math achievement ($\pi_{1,j,k} = -.169, SE = .003, \Delta = .437$), and the contextual effect of L2 math anxiety was also negative ($\beta_{0,0,k} = -.156, SE = .012, \Delta = .411$). Finally, in the total sample of PISA, L1 individual math anxiety was negatively related to math achievement ($\pi_{1,j,k} = -.150, SE = .003, \Delta = .390$), and the contextual effect of L2 math anxiety was also negative ($\beta_{0,0,k} = -.145, SE = .018, \Delta = .380$).

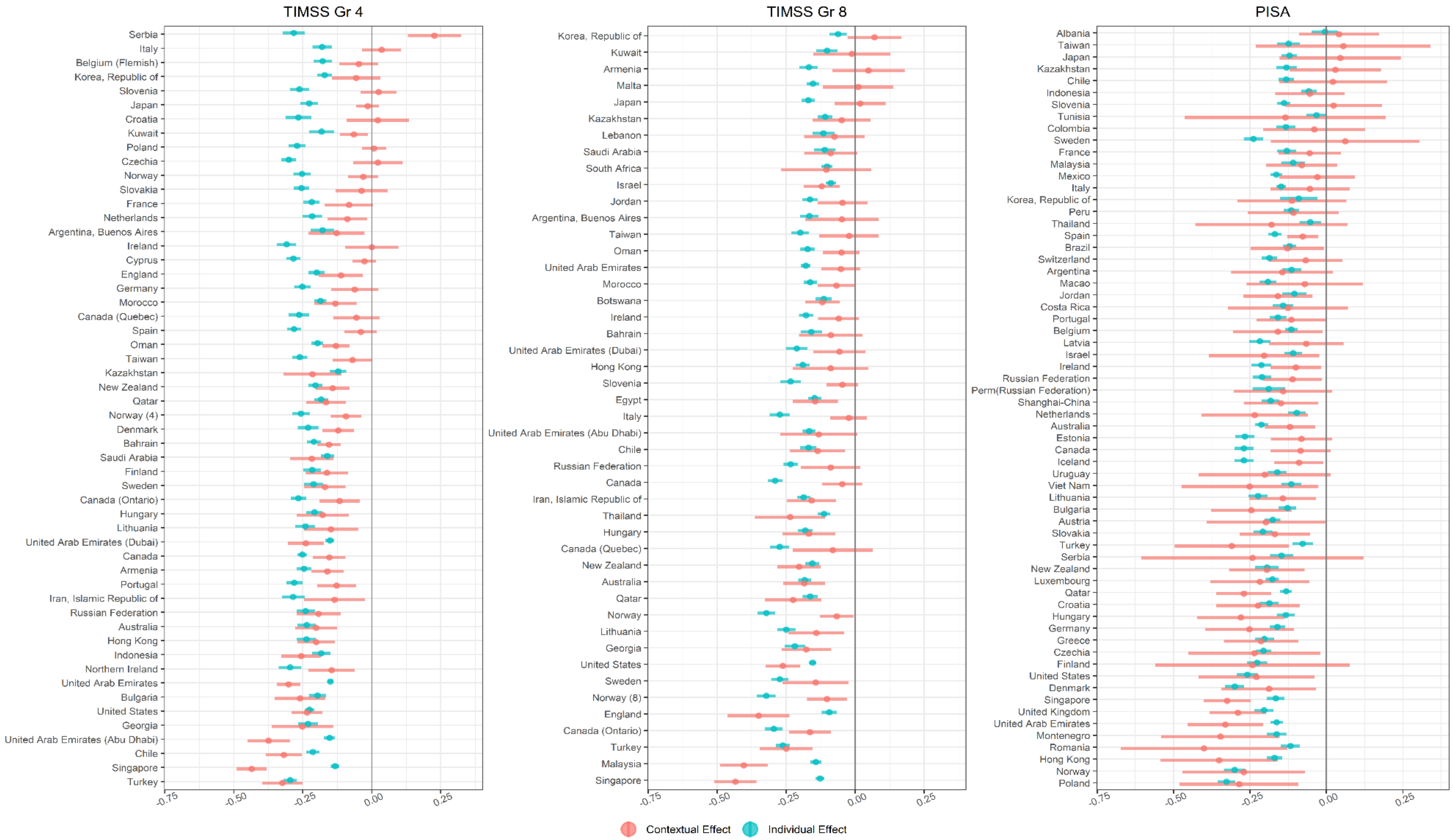
In sum, all three databases yielded negative L1 and L2 associations between math anxiety and math achievement with comparable effect sizes. The negative association between math anxiety and math achievement at L1 suggests that students with higher math anxiety tend to have lower math achievement, while the negative association at L2 suggests that when individual differences in math anxiety is controlled, education environment-average math anxiety has a negative effect on math achievement. The effect size of the individual and contextual effects ranged from .380 to .524, indicating a small to medium effect size (Cohen, 2013).

4.4.2 Are there Between-Country Differences in the Individual and Contextual Effects?

To examine the generalizability of the baseline model, a multigroup analysis was conducted with country membership as the grouping variable. To test whether there are statistically significant between-country differences in the magnitude of the L1 individual effect and L2 contextual effect, a constrained model in which the L1 individual effect and L2 contextual effect are held constant across countries was first estimated for each database. Model fit of the constrained models were poor for all three databases (TIMSS Grade 4: χ^2 = 7878.10, df = 318, CFI = .603, TLI = .863, RMSEA = .056; TIMSS Grade 8: χ^2 = 6987.58, df = 379, CFI = .182, TLI = .724, RMSEA = .048; PISA: χ^2 = 10358.80, df = 270, CFI = .052, TLI = .677, RMSEA = .077), indicating that the restriction of the individual and contextual effects to be equal across countries have significantly reduced model fit.

Given results suggesting that both individual and contextual effects differ between countries, an unconstrained model in which both L1 individual effect and L2 contextual effect are free to vary across countries was next estimated for each database. For the unconstrained model, country specific L1 and L2 parameters are presented in **Figure 7** (see **Appendix 8** for standardized coefficients and effect sizes).

Figure 7. Individual and Contextual Effects of Math Anxiety on Math Achievement.



Note. Countries are ordered according to the average of the two parameters. Confidence Interval (CI) reflect an alpha of 0.05. As such, parameters are statistically significant if the 95% CI does not cross 0 (drawn in black).

For the TIMSS Grade 4, the L1 individual effects of math anxiety on math achievement were statistically significant for all countries and ranged from -.309 to -.122. The L2 contextual effect between math anxiety and math achievement was statistically significant in 37 of the 54 countries and ranged from -.435 to .227. All statistically significant L2 contextual effects were negative except for Serbia, which had a statistically significant positive effect. For the TIMSS Grade 8, L1 individual effect of math anxiety on math achievement were statistically significant for all countries and ranged from -.322 to -.062. The L2 contextual effect of math anxiety on math achievement was statistically significant in 22 of the 46 countries and ranged from -.434 to .007. All statistically significant L2 contextual effects were negative. For the PISA, L1 individual effect of math anxiety on math achievement were statistically significant for all countries except Albania and ranged from -.327 to -.005. The L2 contextual effect of math anxiety on math achievement was statistically significant in 36 of the 64 countries and ranged from -.401 to .061. All statistically significant L2 contextual effects were negative.

In sum, results from the constrained model suggest that there are statistically significant between-country differences in both the individual and contextual effects, and that the relations between the variables from one country cannot be readily generalized to another. Results from the unconstrained model are consistent across databases. Specifically, we observed a robust negative effect between individual math anxiety and math achievement in all countries except one country. However, the L2 contextual effect is more tenuous, with only 47% - 68% of the countries exhibiting a statistically significant negative contextual effect.

Table 11. Results for the Three-Level Model with Math Anxiety as the Outcome Variable

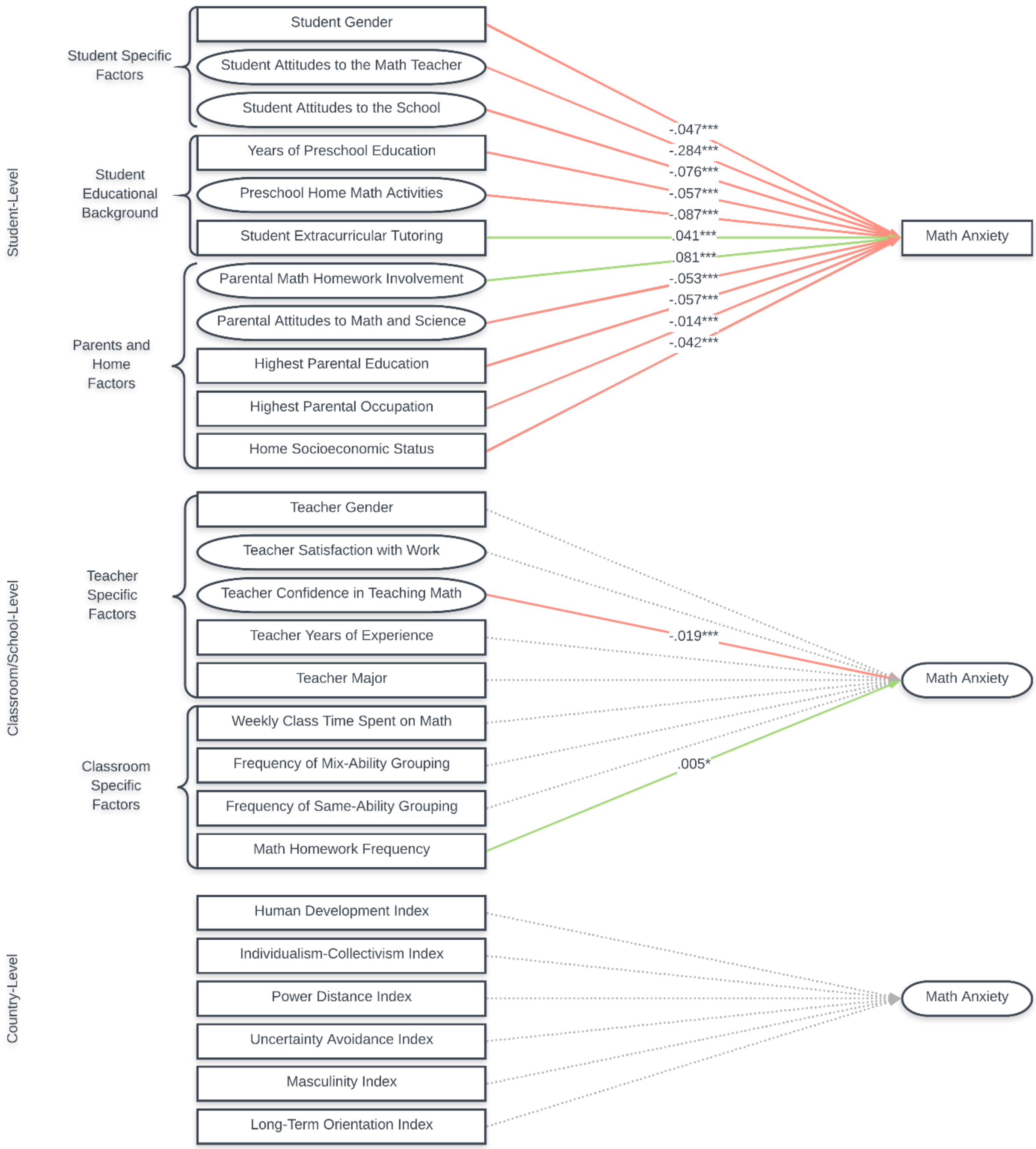
Fixed Effects	*Coefficient(SE)*	*Effect Size(Δ)*
Level 1		
Student Gender, $\pi_{1,j,k}$	-.047(.009)***	.099
Student Attitudes towards the Math Teacher, $\pi_{2,j,k}$	-.284(.029)***	.274
Student Attitudes towards the School, $\pi_{3,j,k}$	-.076(.014)***	.100
Student Years of Preschool Education, $\pi_{4,j,k}$	-.013(.004)***	.027
Student Preschool Home Mathematics Activities, $\pi_{5,j,k}$	-.087(.009)***	.099
Student Current Extracurricular Tutoring/Lessons, $\pi_{6,j,k}$	.041(.007)***	.086
Parental Involvement in Mathematics Homework, $\pi_{7,j,k}$	.081(.009)***	.126
Parental Attitudes towards Mathematics and Science, $\pi_{8,j,k}$	-.053(.007)***	.068
Parents' Highest Education Level, $\pi_{9,j,k}$	-.057(.004)***	.119
Parents' Occupation, $\pi_{10,j,k}$	-.014(.003)***	.029
Home Socioeconomic Status, $\pi_{11,j,k}$	-.042(.005)***	.088
Level 2		
Teacher Gender, $\beta_{0,1,k}$	.011(.008)	.023
Teacher Satisfaction with Work, $\beta_{0,2,k}$	-.005(.005)	.008
Teacher Confidence in Teaching Mathematics, $\beta_{0,3,k}$	-.019(.004)***	.028
Teacher Years of Experience, $\beta_{0,4,k}$	-.006(.004)	.013
Teacher Major, $\beta_{0,5,k}$	-.001(.005)	.002
Weekly Class Time Spent on Mathematics, $\beta_{0,6,k}$	-.001(.003)	.002
Frequency of Mix-ability Grouping, $\beta_{0,7,k}$	.003(.003)	.006
Frequency of Same-ability Grouping, $\beta_{0,8,k}$	.000(.003)	.000
Mathematics Homework Frequency, $\beta_{0,9,k}$	.005(.003)*	.010
Level 3		
Average Initial Math Anxiety, $\gamma_{0,0,0}$	-.013(.028)	-
U.N. Human Development Index, $\gamma_{0,0,1}$	.041(.040)	.086
Individualism-Collectivism, $\gamma_{0,0,2}$	-.084(.057)	.176
Power Distance, $\gamma_{0,0,3}$	.005(.005)	.010
Uncertainty Avoidance, $\gamma_{0,0,4}$	.007(.038)	.015
Masculinity, $\gamma_{0,0,5}$	.052(.048)	.109
Long-Term Orientation, $\gamma_{0,0,6}$	-.065(.040)	.136
Random Effects		
Math Anxiety L1 Residual, $e_{i,j,k}$	.873***	-
Math Anxiety L2 Residual, $r_{0,j,k}$	.034***	-
Math Anxiety L3 Residual, $\mu_{0,0,k}$	.041***	-

Note. *** $p < .001$, ** $p < .01$, * $p < .05$.

4.4.3 What Individual and Environmental Factors Predict Math Anxiety?

To examine the individual and environmental factors that predict math anxiety, we estimated a three-level model with math anxiety as the outcome variable. Due to missing data in the TIMSS Grade 8 and PISA databases, this analysis was solely conducted using the TIMSS Grade 4 database (see methods for more details). Results are presented in **Table 11** and **Figure 8**. Results indicate that a multitude of variables at L1 is associated with math anxiety. The strongest predictor of math anxiety at L1 was students' attitudes towards the math teacher ($\pi_{2,j,k} = -.284(.029), p < .001, \Delta = .274$). Specifically, students' attitudes towards the math teacher's competence and fairness are positively related with lowered math anxiety. At L2, only two variables were significantly related with math anxiety. Specifically, teachers' confidence in teaching math is associated with a reduction in student math anxiety ($\beta_{0,3,k} = -.019(.004), p < .001, \Delta = .028$), and homework frequency is positively associated with increased math anxiety ($\pi_{2,j,k} = -.005(.003), p = 0.048, \Delta = .010$).

Figure 8. Results for the Three-Level Model with Math Anxiety as Outcome.



Note. Rectangles are observed variables, rounded rectangles are latent variables. Red lines refer to negative relations, green lines refer to positive relations.

4.4.4 Do the Predictors of Math Anxiety Serve as Unobserved Confounders to the Relation Between Math Anxiety and Math Achievement?

In the final analysis, the relation between math anxiety and math achievement while controlling for the individual and environmental predictors of math anxiety was explored. To this end, both math anxiety and individual and environmental factors were included as predictors of math achievement. Further, the random-slopes were also regressed onto these predictors as a test of cross-level interactions. Results are presented in **Table 12** and **Figure 9**. Results indicate that after accounting for a large repertoire of potential covariates, the average L1 individual effect ($\gamma_{1,0,0}$) and the average L2 contextual effect ($\gamma_{0,1,0}$) remain robust. The average effect size for the individual effect was .491, and the average effect size for the L2 contextual effect was .312, a small-to-medium effect size.

Further, a cross-level interaction was found in which the greater teacher's confidence in teaching math, the lesser the magnitude of individual math anxiety's effect on math achievement ($\beta_{1,3,k}$). This suggest that teachers with higher confidence in teaching math may be associated with a milder effect of individual math anxiety on math achievement. Similarly, it was found that a country's degree of uncertainty avoidance is related to the degree to which education environment-average math anxiety is related with math achievement ($\gamma_{0,1,4}$). Specifically, countries with higher levels of uncertainty avoidance tend to have education environments in which average math anxiety do not affect student math achievement as severely. Finally, while there are significant variations in the magnitude of the individual effect of math anxiety on math achievement at L3 ($u_{1,0,k}$), none of the country-level predictors significantly predict the variability.

Figure 9. Results for the Three-Level Model with Random Slopes.

Note. Rectangles are observed variables, rounded rectangles are latent variables. S1 refers to the random-slope of individual math anxiety to math achievement, and S2 refers to the random-slope of education environment math anxiety to math achievement. Red lines refer to negative relations, green lines refer to positive relations.

Table 12. Results for the Three-Level Model with Random-Slopes with Covariates for TIMSS Grade 4

Fixed Effects	*Coefficient(SE)*	*Effect Size(Δ)*
Model Math Achievement, π_{0jk}		
Level 1		
Student Gender, π_{2jk}	.028(.004)***	.085
Student Attitudes towards the Math Teacher, $\pi_{3,j,k}$	.039(.017)*	.054
Student Attitudes towards the School, $\pi_{4,j,k}$	.051(.012)***	.096
Student Years of Preschool Education, $\pi_{5,j,k}$	.026(.004)***	.079
Student Preschool Home Math Activities, $\pi_{6,j,k}$	.171(.010)***	.281
Student Current Extracurricular Tutoring/Lessons, $\pi_{7,j,k}$	-.078(.013)***	.236
Parental Involvement in Math Homework, $\pi_{8,j,k}$	-.197(.010)***	.472
Parental Attitudes towards Math and Science, $\pi_{9,j,k}$	.088(.007)***	.164
Parents' Highest Education Level, $\pi_{10,j,k}$	.121(.007)***	.366
Parents' Occupation, $\pi_{11,j,k}$	.049(.004)***	.148
Home Socioeconomic Status, $\pi_{12,j,k}$	.111(.009)***	.336
Level 2		
Teacher Gender, $\beta_{0,2,k}$	-.015(.007)*	.045
Teacher Satisfaction with Work, $\beta_{0,3,k}$	.010(.006)	.023
Teacher Confidence in Teaching Math, $\beta_{0,4,k}$	.029(.006)***	.061
Teacher Years of Experience, $\beta_{0,5,k}$	.010(.008)	.030
Teacher Major, $\beta_{0,6,k}$	.015(.004)***	.048
Weekly Class Time Spent on Math, $\beta_{0,7,k}$	.019(.005)***	.058
Frequency of Mix-Ability Grouping, $\beta_{0,8,k}$	-.008(.004)*	.024
Frequency of Same-Ability Grouping, $\beta_{0,9,k}$	-.015(.004)***	.045
Math Homework Frequency, $\beta_{0,10,k}$	.004(.004)	.012
Level 3		
Average Initial Math Achievement, $\gamma_{0,0,0}$	-.008(.055)	-
Country Math Anxiety, $\gamma_{0,0,1}$	.032(.044)	.097
U.N. Human Development Index, $\gamma_{0,0,2}$	.304(.105)**	.921
Individualism-Collectivism, $\gamma_{0,0,3}$	-.193(.082)*	.585
Power Distance, $\gamma_{0,0,4}$	-.053(.137)	.161
Uncertainty Avoidance, $\gamma_{0,0,5}$	-.066(.077)	.200
Masculinity, $\gamma_{0,0,6}$	.090(.045)*	.273
Long-Term Orientation, $\gamma_{0,0,7}$	.205(.078)**	.621
Model Random Slope of Individual Math Anxiety on Math Achievement, π_{1jk}		
Level 2		
Teacher Gender, $\beta_{1,1,k}$	-.001(.002)	-
Teacher Satisfaction with Work, $\beta_{1,2,k}$	-.001(.003)	-
Teacher Confidence in Teaching Math, $\beta_{1,3,k}$	.006(.003)*	-

Teacher Years of Experience, $\beta_{1,4,k}$	.001(.002)	-
Teacher Major, $\beta_{1,5,k}$	.001(.002)	-
Weekly Class Time Spent on Math, $\beta_{1,6,k}$	.002(.002)	-
Frequency of Mix-Ability Grouping, $\beta_{1,7,k}$	-.002(.002)	-
Frequency of Same-Ability Grouping, $\beta_{1,8,k}$	.002(.002)	-
Math Homework Frequency, $\beta_{1,9,k}$	.001(.002)	-
Level 3		
Individual Math Anxiety Intercept, $\gamma_{1,0,0}$	-.157(.005)***	.476
U.N. Human Development Index, $\gamma_{1,0,1}$	.013(.009)	-
Individualism-Collectivism, $\gamma_{1,0,2}$	-.006(.009)	-
Power Distance, $\gamma_{1,0,3}$	.005(.009)	-
Uncertainty Avoidance, $\gamma_{1,0,4}$	.003(.007)	-
Masculinity, $\gamma_{1,0,5}$	.000(.006)	-
Long-Term Orientation, $\gamma_{1,0,6}$	.010(.005)	-
Model Random Slope of Education Environment-Average Math Anxiety on Math Achievement, π_{1jk}		
Level 3		
Education Environment Math Anxiety Intercept, $\gamma_{0,1,0}$	-.088(.014)***	.267
U.N. Human Development Index, $\gamma_{0,1,1}$	.017(.015)	-
Individualism-Collectivism, $\gamma_{0,1,2}$	.016(.019)	-
Power Distance, $\gamma_{0,1,3}$	-.009(.021)	-
Uncertainty Avoidance, $\gamma_{0,1,4}$	.045(.015)**	-
Masculinity, $\gamma_{0,1,5}$	.001(.011)	-
Long-Term Orientation, $\gamma_{0,1,6}$	.004(.010)	-
Random Effects		
Math Achievement L1 Residual, $e_{i,j,k}$	.374(.012)***	-
Math Achievement L2 Residual, $r_{0,j,k}$	.130(.016)***	-
Math Achievement L3 Residual, $\mu_{0,0,k}$	.109(.027)***	-
Individual Math Anxiety at L2 residual, $r_{1,j,k}$	.001(.000)**	-
Individual Math Anxiety at L3 residual, $u_{1,0,k}$	.001(.000)***	-
Education Environment Math Anxiety at L3 residual, $u_{0,1,k}$	.005(.001)***	-

Note. *** $p < .001$, ** $p < .01$, * $p < .05$.

4.5 Discussion

In recent decades, math anxiety has received increasing research attention, both as a predictor of math achievement, and as phenomenon to be studied and mitigated. Despite this increased attention, the extant literature has remained largely an endeavour of understanding math anxiety as an individual-level phenomenon. However, it is becoming increasingly clear that math anxiety occurs within a complex ecosystem that includes predictors within multiple nested levels. Using multiple large-scale international studies of student achievement, the current study presents the most comprehensive examination of a) the relation between math anxiety and math achievement across the globe, b) the individual and environmental predictors of math anxiety, and c) any potential factors that could explain the relation between math anxiety and math achievement. Multiple important findings were uncovered in the current study.

4.5.1 Summary of Pertinent Findings

Research over the past few decades has revealed multiple predictors of math anxiety that can be found at the level of the individual (e.g., gender (Satake & Amato, 1995), student attitude towards their math teacher (Fast et al., 2010b) etc.) as well as the level of the education environment (e.g., teacher math anxiety (Novak & Tassell, 2017), classroom atmosphere (Radišić et al., 2015) etc.). However, until now, the different levels at which predictors of math anxiety can be identified have largely been studied in isolation of one another. To the best of our knowledge, the current study represents the largest and most culturally diverse study on math anxiety. Importantly, the analyses explicitly model the fact that math anxiety is accounted for by factors that occur at multiple nested levels of analysis – that students are nested within schools which are themselves nested within countries.

The findings from this multilevel, cross-national study of math anxiety reveal that the most consistent predictors of mathematics anxiety can be found at the individual level. In other words, factors that are unique to individual students, independently of the country or educational environment that they are situated within, are the most consistent predictors of the level of math anxiety that they experience. One of the factors that stands out at the individual level is the confidence that students have in their teacher. While effects of the education environment level were less consistent, teacher confidence in teaching math was negatively associated with students' math anxiety. Furthermore, the frequency with which math homework was assigned with in the education environment that students are in also contributed to math anxiety. Interestingly, the above reported analyses do not reveal any predictors of math anxiety at the country level.

With respect to the relationship between math anxiety and math achievement, the present analyses reveal a consistent relationship between individual math anxiety and math achievement across the globe. Moreover, this effect is nuanced by the fact that at least in some countries, the average math anxiety experienced by students with in the same educational context (i.e., school or classroom) is related to the individual students' math achievement, independently of the individual level of math anxiety that students report. This means then when considering how math anxiety exerts its effect on math achievement, it is important to take into account the effects of the environment within which the student is learning math.

Importantly, we also found evidence of cross-level interactions – where the strength of the individual and contextual effect of math anxiety are dependent on variables attributable to the larger environment in which these effects are nested. First, we found evidence to suggest that teacher confidence in teaching math is associated with a smaller effect of an individual students' math anxiety on their math achievement. Second, we found that in countries with higher levels of uncertainty avoidance, the effects of the education environment average math anxiety on student math achievement, is lower.

4.5.2 Individual and Contextual Effect of Math Anxiety on Math Achievement

While the current study employed databases that differ with respect to age groups, countries sampled, and method of measuring math anxiety and math achievement, there nevertheless was a surprising level of concordance in the results regarding the individual and contextual effect of math anxiety on math achievement. At the individual level, our results from all three databases suggest that students in virtually all countries have exhibited a negative association between individual math anxiety and math achievement. At the education environment level, results suggest that approximately half the countries exhibited a statistically significant association between education environment average math anxiety and math achievement.

The consistency of the individual effect of math anxiety across databases and countries is concordant with the idea that math anxiety is a "global phenomenon" (Foley et al., 2017, p.52), and highlights that universality of the adverse effects of individual math anxiety on math performance. Further, we have revealed a novel, contextual effect of math anxiety, whereby the average level of student math anxiety in one's immediate education environment makes an independent contribution to explaining variability in math achievement. However, results suggest that that the contextual effect is highly variable across countries, with only half of the sampled countries exhibiting a statistically significant contextual effect. Consequently, this suggests there is heterogeneity in the mechanisms through which education environment average math anxiety affects math achievement and suggest that studies examining the interactions between these variables may have low generalizability across countries.

4.5.3 Individual and Environmental Predictors of Math Anxiety

Our results indicate that math anxiety is associated with variety of individual and environmental factors. Our findings reinforce research findings from multiple previous studies. For example, that there are gender differences in math anxiety (Griggs et al., 2013; Satake & Amato, 1995; Yüksel-Şahin, 2008), and that students' own attitudes towards the learning environment to be negatively associated with math anxiety (e.g., Fast et al., 2010; Radišić et al., 2015). Given that virtually all the independent variables at the individual level were associated with math anxiety, our findings support the notion that math anxiety is a multifaceted phenomenon.

At the education environment level, we found far fewer variables associated with math anxiety, and generally, with lower effect sizes. However, the variables that were found to be associated with math anxiety align with prior findings, providing an additional weight of evidence. Specifically, that teacher's confidence in math teaching – which is negatively correlated with teacher math anxiety (Bursal & Paznokas, 2006; Geist, 2015) – is related with lower student math anxiety (Novak & Tassell, 2017; Ramirez et al., 2018), and the frequency of homework being associated with higher math anxiety (Cheema & Sheridan, 2015; Radišić et al., 2015).

Taken together, our results consolidate current understanding of the causes of math anxiety and supports the notion that math anxiety is affected by multiple factors. Further, our results also suggest that, in contrast to predictors at the individual level, correlates of math anxiety at the education environment level are more specific and are limited to only a few factors.

4.5.4 Individual and Contextual Effects in the Context of other Predictors

When considering math anxiety as a predictor of math achievement in the context of other potential predictors of math achievement, we find that the average individual and contextual effect of math anxiety remain strong predictors of math achievement. Interestingly, two cross-level interactions were found. First, that higher teacher confidence in teaching math was related with a weaker relation between individual math anxiety and math achievement, and that higher country uncertainty avoidance is associated with a lower contextual effect of math anxiety. While the effects are generally small, these results are preliminary evidence to suggest that variables at higher levels not only could simply predict student math anxiety, but also could modulate the individual and contextual effects of math anxiety.

4.5.5 Teachers Playing a Critical Role

In the current study, we found evidence to suggest that teachers may play a central role in the effects of math anxiety. Specifically, we find a) students' perception of teacher competence to be the strongest predictor of student math anxiety at the individual level, b) negative associations between teacher confidence in teaching math and education environment average math anxiety, and c) teacher confidence in teaching math also being associated with a weaker relationship between individual math anxiety and math achievement.

Importantly, extant studies suggest that teacher confidence and teacher math anxiety are negatively correlated (Bursal & Paznokas, 2006; Geist, 2015), and it is unclear whether the association between teacher confidence and math achievement is due to this negative correlation. Future studies will be required to identify whether one or both teacher confidence and teacher math anxiety are related with math achievement.

4.5.6 The Effects of Homework

Interestingly, our results suggest that homework may play a significant role in math anxiety. In line with previous findings, we find that the frequency of homework assigned is related with higher math anxiety (Cheema & Sheridan, 2015; Radišić et al., 2015). Similarly, we also find evidence to suggest that parental involvement with homework is associated with an increase in math anxiety (Maloney et al., 2015). These findings suggest that homework and the degree to which parents are involved in their children's homework must be considered carefully in any future study of math anxiety as well as potential interventions to alleviate math anxiety.

4.6 Conclusion

As a whole, the present data reveal that mathematical anxiety and its relationship to math achievement is affected by factors that are unique to the individual child, the educational context within which they learn, and one's country of residence. These data highlight the importance of moving beyond positioning math anxiety as something that exists only within an individual student but as a construct that is affected, in complex ways, by factors that are nested within the educational environment and the country of the learner.

4.7 References

Arnett, J. (2008). The weirdest people in the world. American Psychologist, 63(7), 602–614.

Ashcraft, M. H., Donley, R. D., Halas, M. A., & Vakali, M. (1992). Working memory, automaticity, and problem difficulty. In Advances in psychology (Vol. 91, pp. 301–329). Elsevier.

Ashcraft, M. H., & Faust, M. W. (1994). Mathematics anxiety and mental arithmetic performance: An exploratory investigation. Cognition & Emotion, 8(2), 97–125.

Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. Psychonomic Bulletin & Review, 14(2), 243–248.

Ashcraft, M. H., & Ridley, K. S. (2005). Math anxiety and its cognitive consequences: A tutorial review.

Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2020). A meta-analysis of the relation between math anxiety and math achievement. Psychological Bulletin.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. Proceedings of the National Academy of Sciences, 107(5), 1860–1863.

Berkowitz, T., Schaeffer, M. W., Maloney, E. A., Peterson, L., Gregor, C., Levine, S. C., & Beilock, S. L. (2015). Math at home adds up to achievement in school. Science, 350(6257), 196–198.

Brincks, A. M., Enders, C. K., Llabre, M. M., Bulotsky-Shearer, R. J., Prado, G., & Feaster, D. J. (2017). Centering predictor variables in three-level contextual models. Multivariate Behavioral Research, 52(2), 149–163.

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. Acta Psychologica, 148, 163–172. https://doi.org/10.1016/j.actpsy.2014.01.016

Chipman, S. F., Krantz, D. H., & Silver, R. (1992). Mathematics anxiety and science careers among able college women. Psychological Science, 3(5), 292–296.

Cohen, J. (2013). Statistical power analysis for the behavioral sciences. Academic press.

Diez-Roux, A. V. (1998). Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. American Journal of Public Health, 88(2), 216–222.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. Psychological Methods, 12(2), 121.

Fast, L. A., Lewis, J. L., Bryant, M. J., Bocian, K. A., Cardullo, R. A., Rettig, M., & Hammond, K. A. (2010). Does math self-efficacy mediate the effect of the perceived classroom environment on standardized math test performance? Journal of Educational Psychology, 102(3), 729.

Faust, M. W. (1996). Mathematics anxiety effects in simple and complex addition. Mathematical Cognition, 2(1), 25–62.

Foley, A. E., Herts, J. B., Borgonovi, F., Guerriero, S., Levine, S. C., & Beilock, S. L. (2017). The math anxiety-performance link: A global phenomenon. Current Directions in Psychological Science, 26(1), 52–58.

Graham, J. W., & Hofer, S. M. (2000). Multiple imputation in multivariate research.

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. Journal for Research in Mathematics Education, 33–46.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? Behavioral and Brain Sciences, 33(2–3), 61–83.

Hofstede, G. H., Hofstede, G. J., & Minkov, M. (2010). Cultures and organizations: Software of the mind (3rd ed.). Mcgraw-hill New York.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1–55. https://doi.org/10.1080/10705519909540118

Hutchison, D., & Schagen, I. (2007). Comparisons between PISA and TIMSS: Are we the man with two watches. Lessons Learned: What International Assessments Tell Us about Math Achievement, 227–261.

Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. Structural Equation Modeling, 8(3), 325–352.

Kreft, I. G., & Yoon, B. (1994). Are multilevel techniques necessary? An attempt at demystification.

Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. Learning and Individual Differences, 19(3), 355–365.

Lyons, I. M., & Beilock, S. L. (2011). Mathematics anxiety: Separating the math from the anxiety. Cerebral Cortex, 22(9), 2102–2110.

Lyons, I. M., & Beilock, S. L. (2012). When math hurts: Math anxiety predicts pain network activation in anticipation of doing math. PloS One, 7(10), e48076.

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. Journal for Research in Mathematics Education, 520–540.

Maloney, E. A., Ansari, D., & Fugelsang, J. A. (2011). Rapid Communication: The effect of mathematics anxiety on the processing of numerical magnitude. Quarterly Journal of Experimental Psychology, 64(1), 10–16.

Maloney, E. A., Waechter, S., Risko, E. F., & Fugelsang, J. A. (2012). Reducing the sex difference in math anxiety: The role of spatial processing ability. Learning and Individual Differences, 22(3), 380–384.

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. Educational Psychologist, 47(2), 106–124.

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. Multivariate Behavioral Research, 44(6), 764–802.

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K.-T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish–little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. Educational Psychology Review, 20(3), 319–350.

Martin, M. O., Mullis, I. V., & Hooper, M. (2016). Methods and procedures in TIMSS 2015. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA).

Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. Journal of Educational Psychology, 82(1), 60.

Mizala, A., Martínez, F., & Martínez, S. (2015). Pre-service elementary school teachers' expectations about student performance: How their beliefs are affected by their mathematics anxiety and student's gender. Teaching and Teacher Education, 50, 70–78.

Muthén, L. K., & Muthén, B. O. (2010). Mplus: Statistical analysis with latent variables: User's guide. Muthén & Muthén Los Angeles.

OECD. (2014a). PISA 2012 technical report. OECD publishing Paris.

OECD, P. (2014b). Results: Creative Problem Solving: Students' Skills in Tackling Real-Life Problems (Volume V), PISA. OECD publishing.

Pletzer, B., Kronbichler, M., Nuerk, H.-C., & Kerschbaum, H. H. (2015). Mathematics anxiety reduces default mode network deactivation in response to numerical tasks. Frontiers in Human Neuroscience, 9, 202.

Ramirez, G., Chang, H., Maloney, E. A., Levine, S. C., & Beilock, S. L. (2016). On the relationship between math anxiety and math achievement in early elementary school: The role of problem solving strategies. Journal of Experimental Child Psychology, 141, 83–100.

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. Journal of Cognition and Development, 14(2), 187–202.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (Vol. 1). Sage.

Rubinsten, O., Eidlin, H., Wohl, H., & Akibli, O. (2015). Attentional bias in math anxiety. Frontiers in Psychology, 6, 1539.

Sarkar, A., Dowker, A., & Kadosh, R. C. (2014). Cognitive enhancement or cognitive cost: Trait-specific outcomes of brain stimulation in the case of mathematics anxiety. Journal of Neuroscience, 34(50), 16605–16610.

Schaeffer, M. W., Rozek, C. S., Maloney, E. A., Berkowitz, T., Levine, S. C., & Beilock, S. L. (2020). Elementary school teachers' math anxiety and students' math learning: A large‐scale replication. Developmental Science, e13080.

Schafer, J. L. (1999). Multiple imputation: A primer. Statistical Methods in Medical Research, 8(1), 3–15.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. Review of Educational Research, 75(3), 417–453.

Soni, A., & Kumari, S. (2017). The role of parental math anxiety and math attitude in their children's math achievement. International Journal of Science and Mathematics Education, 15(2), 331–347.

Swars, S. L., Daane, C. J., & Giesen, J. (2006). Mathematics anxiety and mathematics teacher efficacy: What is the relationship in elementary preservice teachers? School Science and Mathematics, 106(7), 306–315.

United Nations Development Program. (2015). Human Development Data (1990-2018) [Database]. Human Development Data (1990-2018). http://hdr.undp.org/en/data

Van den Noortgate, W., Opdenakker, M.-C., & Onghena, P. (2005). The effects of ignoring a level in multilevel analysis. School Effectiveness and School Improvement, 16(3), 281–303.

Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful. IERI Monograph Series, 2(1), 9–36.

Vukovic, R. K., Kieffer, M. J., Bailey, S. P., & Harari, R. R. (2013). Mathematics anxiety in young children: Concurrent and longitudinal associations with mathematical performance. Contemporary Educational Psychology, 38(1), 1–10.

Wang, Z., Hart, S. A., Kovas, Y., Lukowski, S., Soden, B., Thompson, L. A., Plomin, R., McLoughlin, G., Bartlett, C. W., & Lyons, I. M. (2014). Who is afraid of math? Two sources of genetic variance for mathematical anxiety. Journal of Child Psychology and Psychiatry, 55(9), 1056–1064.

Chapter 5

5.1 General Discussion

The broad goal of the current thesis was to contribute to the ongoing and expanding research on how children develop numerical and mathematical skills and understanding. Study 1 and study 2 focused on the ongoing debate regarding how children initially acquire the uniquely human ability to represent quantities with symbolic numbers, and how this foundational skill may affect math achievement. Study 3 examined the emerging issue of how math anxiety may be related to student achievement. Through the use of structural equation modelling, the three studies have yielded novel insights, that provide important new insights into children's development of numerical and mathematical skills.

5.2 Study 1 and Study 2 — The Relation between Numerical Development and Math Achievement

One of the key mathematical skills that children need to master in early childhood is the skill to understand the meaning of numerical symbols (e.g. that spoken words, Arabic numerals refer to specific quantities). Children's understanding of numerical symbols is thought to be a foundational skill that predicts children's later math achievement (Feigenson et al., 2004). So how do children learn the meaning of numerical symbols? Traditionally, it has been hypothesized that the learning of symbolic numbers may be related with an innate cognitive system to perceive quantities (the approximate number system; ANS) that children possess (Feigenson et al., 2004). Specifically, under this mapping account, it has been hypothesized that children acquire this ability via a one-to-one mapping between symbolic numerals and their corresponding non-symbolic representation in the ANS (Dehaene, 2007).

An impressive corpus of research evidence has been uncovered in support of the mapping account; however, recent studies have uncovered results that cannot be easily explained by the mapping account. Indeed, one major debate in the current literature pertains to one of the central hypotheses posited by the mapping account – that earlier ANS ability predicts later math achievement. Specifically, while some studies have found that ANS ability relates to math achievement (e.g., DeWind & Brannon, 2012; Halberda & Feigenson, 2008), others have failed to find such a relation (e.g., Holloway & Ansari, 2008; Sasanguie et al., 2013).

In response to these controversial findings, an alternative, so called, refinement account has been proposed as a potentially more accurate description of children's developing understanding of the meaning of numerical symbols. In the refinement account, it is hypothesized that once a basic understanding of symbolic numbers is established, the repeated practice and usage of numerical symbols are self-reenforcing. Symbolic representations initially based on the object tracking system will be increasingly overwritten and reinterpreted by their relations with other numerical symbols (Mix et al., 2002; Nieder, 2009).

Against this background, study 1 and study 2 addressed two outstanding questions in the extant literature. First, whether there are longitudinal patterns of development that would support one developmental account over another. Second, whether the two developmental accounts adequately describe the developmental trajectory of the whole population.

5.2.1 Which of the Two Developmental Accounts Better Fit the Data?

In study 1, we tested 4 Random-Intercept Cross-Lagged Panel Analysis (RI-CLPMs) reflecting the mapping account, refinement account and explored the possibility of bidirectional influences between ANS ability and symbolic number ability. The four RI-CLPM contrasts multiple hypotheses posited by the mapping and refinement accounts.

According to the mapping account, the ANS plays a central role in children's development. First, it is proposed that symbolic numerals are mapped on to the ANS (Dehaene, 2005), and therefore earlier ANS ability predict later symbolic number ability. Second, as symbolic numbers and non-symbolic numbers draw on the same underlying representations, children with stronger ANS ability will have an easier time retrieving corresponding symbol attached to any non-symbolic number stimuli (Geary, 2013). As such, earlier ANS ability predict later translating ability. Finally, ANS ability has been posited to be directly related with math achievement (Dehaene, 2005; Feigenson et al., 2004; Gilmore et al., 2010), or to be indirectly related with math achievement with symbolic number ability being the mediator (Price & Fuchs, 2016).

According to the refinement account, the continual usage and practice of symbolic numbers catalyze changes in the underlying neural circuitries responsible for magnitude processing, leading to a refinement of the ANS (Verguts & Fias, 2004). Therefore, earlier symbolic number ability predicts later ANS ability. Second, the process of practice and usage of symbolic numbers stimulate the initial understanding of symbolic numerals based on the object tracking system to be overwritten by their relation with other symbolic numbers, (Mix et al., 2002; Nieder, 2009), the efficiency in translating between numerical formats would be contingent on the rate at which the semantic number understanding matures (Lyons et al., 2018). As such, earlier symbolic number ability predicts later ANS ability. Finally, children with more familiarity and practice with symbolic numbers will be better positioned to learn to manipulate these symbols. Consequently, earlier symbolic number ability predicts later math achievement.

Results are highly supportive of the refinement account. Specifically, results suggest that symbolic number ability was a strong predictor of later ANS ability, translating ability and math achievement, both from T1 to T2, and T2 to T3. Additionally, transient effects, whereby translating ability predicts later symbolic number ability and math achievement, was also observed. Taken together, these results suggest that the refinement account best fits observed developmental data. Further, our results also suggest that translating ability may have complex or transient effects, and future studies will be required to ascertain whether these effects exist and the reasons behind them.

5.2.2 Subgroups in the Population

Study 2 examines the degree to which the two abovementioned developmental accounts can adequately describes children's development. One the one hand, both the mapping and refinement accounts of numerical development assumes the population is develops homogenously. That is, all children would go through the same developmental process in learning symbolic numbers. While on the other hand, emergent research provides evidence that there exists substantial heterogeneity in the developmental trajectories that children undergo (Bartelet et al., 2014; Chew et al., 2016, 2019; de Souza Salvador et al., 2019; Kißler et al., 2021).

Study 2 employed a Random-Intercept Latent Transition Analysis model that removes the influences of between-person influences and ascertains whether there are patterns of performance that can be grouped in meaningful ways. Results suggests that four profiles can be reliably extracted. Examination of the profile means and the transition probabilities suggests that these four profiles best reflect levels of development. Indeed, our results suggests that the population is quite homogenous, with 96% of all students having seen some degree of progression through these developmental levels. These results speak against the notion that there are different subgroups in the population that exhibits different developmental profiles. Instead, our results support the notion that the hypothesized unidirectional associations between the ANS ability, symbolic number ability and math achievement are applicable to all members of the population. These results support study 1, and suggest that not only the refinement account best describes data, but also that most children undergo similar developmental processes as described by the refinement account.

However, our results also suggest that while a majority of students (56%) see consistent progression through these developmental levels, a significant proportion of students (40%) of students seeing inconsistent progression or no progression at all. This suggests that there is heterogeneity regarding the pace of progression. Consequently, this suggest that studies with smaller sample sizes may yield more inconsistent estimates of association between the variables, as it is easier for smaller samples to contain a majority of students with inconsistent progression or no progression by chance, leading to a smaller estimate of association. As such, our results suggests that one potential reason for the divergence in research findings may be the heterogeneity in how quickly children progress through developmental levels.

Finally, we examined whether students' Educational Quality and Accountability Office Assessment (EQAO) Scores (collected in Grade 3) are related with 1) the random intercept – an estimate of student's time-invariant, trait-like variability (e.g., in intelligence, SES, interest in mathematics etc.) and 2) profile membership at the end of Grade 1. Results suggest that children's EQAO scores are significantly predicted by the random-intercept, and there are significant mean differences in EQAO scores depending on profile membership. Examination of changes in EQAO scores suggest that both between-person variability (as measured by the random-intercept) and within-person variability (as measured by the level of development) predict EQAO scores similar in magnitude.

In sum, study 2 revealed that a majority of the population go through the same developmental trajectory, but the pace of development is highly varied.

5.3 Study 3 – Math Anxiety and Math Achievement

In recent decades, increasing research focus has been placed on math anxiety, both as a predictor of math achievement, and as an outcome to be studied. However, the extant literature has been marked by information silos that have left multiple research questions overlooked. First, while studies much of the studies published to-date examining math anxiety's effect on math achievement as a within-individual effect, and few studies have examined whether there is a contextual effect of math anxiety (e.g., whether one peer's math anxiety may affect math achievement). Second, most current research on math anxiety have been done in Western, Educated, Industrialized, Rich and Democratic countries (Henrich et al., 2010), and little is known regarding the degree to which current research findings can generalize to other cultures. Using three international studies of achievement and a multilevel modelling approach, study 3 addressed these three research questions and have arrived at several interesting and important results.

5.3.1 The Contextual Effect of Math Anxiety and Between Country Differences

Study 3 first examined whether the level of math anxiety of one's peers in an education environment (i.e., average education environment math anxiety) may predict student math anxiety over and above what could be predicted by individual math anxiety – a contextual effect. To this end, study 3 estimated both the individual effect and contextual effect for each country. Results suggest that there is a high degree of homogeneity in the individual effect – with virtually all countries exhibiting an individual effect of math anxiety. Further, results indicate there while there is evidence for a contextual effect of math anxiety, there is considerable heterogeneity with regards to the magnitude of effect. Specifically, only approximately half of the countries exhibited a statistically significant contextual effect.

Taken together, these results suggest that the degree to which students' own math anxiety affect performance is relatively similar between countries, but the degree to which peers' math anxiety affect performance is more variable and dependent on country membership. Consequently, these results suggest that between country differences in the contextual effect may be the main driver of the observed between country differences in math anxiety's effects. For instance, our results suggest that the reason behind the higher correlation between math anxiety and achievement in East Asian countries (e.g., Foley et al., 2017) may be due to these countries having a higher contextual effect when compared with other countries.

5.3.2 Reasons for Heterogeneity in the Individual and Contextual Effect

When environmental covariates (e.g., teacher confidence in teaching math, home support, SES, student attitudes) are added as predictors of math anxiety, results indicate that both the average individual effect and average contextual effect remain statistically significant. Importantly, when these environmental covariates are used to predict the variability in the individual and contextual effects, two cross-level interactions were observed. First, it was found that a teacher's level of confidence in teaching mathematics is associated with a reduction in the individual effect. In other words, the adverse effects of students' own math anxiety on math achievement are reduced when teachers are more confident in their own ability. Critically, this association seems to be independent of other indicators of teacher competence – such as teacher years of experience or teacher major. Taken together, this suggests that teacher's affective tendencies may be associated with the severity of the individual effect.

Second, country's level of uncertainty avoidance – the degree to which the members of a culture feel threatened by ambiguous or unknown situations – is related with a reduction in the contextual effect. One possible reason for this association is that for countries with higher uncertainty avoidance, class activities, homework and testing is more predictable, while for countries with lower uncertainty avoidance, these factors are less predictable. Higher unpredictability of these class factors may lead to more situations in which students feel they are underprepared to handle, increasing math anxiety.

5.4 The Ubiquity of Multilevel Models

One underlying theme of all three studies is that the multilevel nature of the data necessitates more advanced statistical techniques to arrive at the correct statistical inference. For example, the statistical models used in study 1 and study 2 included random-intercept elements that separate within-person variability (i.e., instances in time) and between-person variability. Similarly, in study 3, independent influences of the individual, education environment, and country were separately estimated. In addition to demonstrating that data in education is often multilevel in nature, the three studies also demonstrated that ignoring the nested nature of the data may lead researchers to misinterpret or overlook relations.

In study 1 and study 2, we pointed out that extant research findings are often mismatched with theoretical hypotheses. Specifically, extant research often offers between-individual correlations between the variables, but theoretical hypotheses posits within-individual associations (Carey, 2009; Curran & Bauer, 2011; Feigenson et al., 2004). This mismatching is predicated on the assumption that between-person relations among variables reflect the aggregate of within-person relations (Diez-Roux, 1998). Critically, as it is possible for between-person and within-person relations between variables to differ, failing to examine this assumption may lead researchers to misleading conclusions.

Further, due to the included random intercepts in study 1 and study 2, research findings in both studies are more nuanced and specific. This is because the random intercept removes all between-person variability in the data, leaving inferences regarding longitudinal relations among variables (study 1) and performance profiles (study 2) to be solely predicated on within-person variability. This increased specificity allows for more certainty in the interpretation of results.

In study 1, longitudinal associations between variables (e.g., earlier symbolic number ability predicting ANS ability) can be interpreted as within-person associations clear of any influences by time-invariant, state-like covariates (e.g., intelligence, SES, interest in math etc.). This "purer" estimate of longitudinal association would be more difficult and costly to estimate using more traditional methods (e.g., collecting all potential time-invariant, state-like covariates and adding them into the model). Similarly, in study 2, the estimated profiles are similarly free of between-person variability (e.g., differences in intelligence, SES etc.). Therefore, the estimated profiles would better reflect qualitative differences in transient states (i.e., levels of development), rather than phenotypical traits difference between groups in the population (e.g., low-performing vs. high-performing).

In study 3, in contrast to the majority of extant research examining the relation between math anxiety and math achievement as a within-person process. We revealed that there exists a contextual math anxiety (i.e., the level of math anxiety in one's peer group or in one's culture group) which has an independent effect on student's math achievement. This overlooked relation among the variables leaves current understanding of math anxiety incomplete.

Distinguishing between the individual and contextual effect of math anxiety is more nuanced, as it shows that it important to not only account for students' individual math anxiety, but also account for the level of math anxiety in students' immediate learning environment. Further, these findings also open new avenues of inquiry for future studies. For instance, in study 3, we examined between-country differences in the individual and contextual effects and the reasons behind these between-country differences.

Taken together, studies 1 to 3 have demonstrated that since data regarding children's development is often multilevel in nature, disregarding this nested structure may lead to misleading conclusions or overlooked relations. Therefore, using the appropriate statistical analyses lead to more nuanced results and opens new avenues of inquiry.

5.5 Conclusion

In summary, using three separate studies, the current thesis examined two hotly debated issues in the field of numerical and mathematical cognition. Study 1 and study 2's examination of the mapping and refinement accounts reveal that the refinement account best describes students' development and that most students in the population exhibit the same developmental trajectory. Study 3's examination of math anxiety has uncovered a novel contextual effect of math anxiety, suggesting that peer's math anxiety may predict one's math achievement. Together, these studies have contributed to clarifying and expanding current understanding of the cognitive and affective antecedents of math achievement, and have opened new avenues of inquiry futures studies could undertake.

5.6 References

Bartelet, D., Ansari, D., Vaessen, A., & Blomert, L. (2014). Cognitive subtypes of mathematics learning difficulties in primary education. Research in Developmental Disabilities, 35(3), 657–670.

Carey, S. (2009). The Origin of Concepts. Oxford University Press.

Chew, C. S., Forte, J. D., & Reeve, R. A. (2016). Cognitive factors affecting children's nonsymbolic and symbolic magnitude judgment abilities: A latent profile analysis. Journal of Experimental Child Psychology, 152, 173–191.

Chew, C. S., Forte, J. D., & Reeve, R. A. (2019). Implications of Change/Stability Patterns in Children's Non-symbolic and Symbolic Magnitude Judgment Abilities Over One Year: A Latent Transition Analysis. Frontiers in Psychology, 10, 441.

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. Annual Review of Psychology, 62, 583–619.

de Souza Salvador, L., Moura, R., Wood, G., & Haase, V. G. (2019). Cognitive heterogeneity of math difficulties: A bottom-up classification approach. Journal of Numerical Cognition, 5(1), 55–85.

Dehaene, S. (2005). Evolution of human cortical circuits for reading and arithmetic: The "neuronal recycling" hypothesis. From Monkey Brain to Human Brain, 133–157.

Dehaene, S. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In Sensorimotor foundations of higher cognition (pp. 527–574).

DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: Effects of feedback and training. Frontiers in Human Neuroscience, 6. https://doi.org/10.3389/fnhum.2012.00068

Diez-Roux, A. V. (1998). Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. American Journal of Public Health, 88(2), 216–222.

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

Foley, A. E., Herts, J. B., Borgonovi, F., Guerriero, S., Levine, S. C., & Beilock, S. L. (2017). The math anxiety-performance link: A global phenomenon. Current Directions in Psychological Science, 26(1), 52–58.

Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. Current Directions in Psychological Science, 22(1), 23–27.

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.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the 'number sense': The approximate number system. In 3-, 4-, 5-, 6-Year-Olds and Adults. Developmental Psychology, 1457–1465. https://doi.org/10.1037/a0012682

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? Behavioral and Brain Sciences, 33(2–3), 61–83.

Holloway, I. D., & Ansari, D. (2008). Domain‐specific and domain‐general changes in children's development of number comparison. Developmental Science, 11(5), 644−649.

Kißler, C., Schwenk, C., & Kuhn, J.-T. (2021). Two dyscalculia subtypes with similar, low comorbidity profiles: A mixture model analysis. Frontiers in Psychology, 12, 1828.

Lyons, I. M., Bugden, S., Zheng, S., De Jesus, S., & Ansari, D. (2018). Symbolic number skills predict growth in nonsymbolic number skills in kindergarteners. Developmental Psychology, 54(3), 440. https://doi.org/10.1037/dev0000445

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). Quantitative development in infancy and early childhood. Oxford University Press.

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

Price, G. R., & Fuchs, L. S. (2016). The Mediating Relation between Symbolic and Nonsymbolic Foundations of Math Competence. PLOS ONE, 11(2), e0148981. https://doi.org/10.1371/journal.pone.0148981

Sasanguie, D., Göbel, S., Moll, K., Smets, K., & Reynvoet, B. (2013). Acuity of the approximate number sense, symbolic number comparison or mapping numbers onto space: What underlies mathematics achievement? Journal of Experimental Child Psychology, 114(3), 418–431.

Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: A neural model. Journal of Cognitive Neuroscience, 16(9), 1493–1504. https://doi.org/10.1162/0898929042568497

Appendices

Appendix 1 - Discrepancies between Preregistered Analysis and Reported Analysis

Preregistered Analysis	**Reported Analysis**	**Justification**
- Dummy coded covariates were preregistered to be used for all comparison and verification tasks for T1 – T3.	- Dummy coded covariates were used for only T1 and T2. T3 dummy coded covariates were not included for all tasks.	- Due to only a few participants having negative scores in T3 (n < 20), the inclusion of dummy coding causes instability in model fit with insignificant changes to the relationship between variables of interest. This was discovered when performing outlier analysis: removal of participants with T3 negative scores causes substantial changes in model fit. However, the removal of the dummy variable eliminated the changes in model fit. This suggests the substantial changes in model fit only relates to regression coefficients relating the dummy variable to the variables of interest. As such, the decision to remove all T3 dummy variables was made. The inclusion or exclusion of T3 dummy variables did not significantly change the interpretation of results.
- Multivariate outliers were planned to be detected using the	- Model fit outlier detection and removal followed the guidelines laid out by	- Due to computational difficulties associated with missing data when calculating the MCD distance, we

Mahalanobis – Minimum covariance determinant (MCD) distance at a breakdown point of 75%.	Aguinis, Gottfredson, & Joo (2013).	opted to follow another guideline to remove outliers.
- No data rescaling was preregistered prior to analysis.	- Data were rescaled by multiplying or dividing continuous variables by a constant so as to restrict variance to be between 1 – 10.	- Convergence issues were encountered in Mplus due to vastly different scales between continuous variables in the dataset. It was recommended by Mplus support that continuous variables to be rescaled to have variance between 1 and 10 to avoid these issues. Such rescaling does not affect relationships between variables (Little, 2013).
- Grade 1 Term 2 academic achievement was operationalized as the sum scores of the sub-scales in report card grades.	- Grade 1 term 2 academic achievements were operationalized as the mean score of the sub-scales in the report card grades if at least 3 of the 5 sub-scales were not missing. Else, the data point was treated as missing data.	- The decision to use either sum score or mean scores would have no effect on results in the absence of missing data. Due to the presence of missing data, the sum score would skew the results of participants with partially missing data.
- Figure 2 (bi-directional/time-dependent model) illustrated in the	- DC > MC and MC > DC were included.	- This was an illustration error in the preregistration that was corrected in the final report. These regressions

preregistration document did not associate DC > MC, and MC > DC for all time-points.		can be inferred by reading the text of the document.
- No mention of assessment of normality was preregistered.	- Normality was qualitatively assessed.	- This was an oversight when compiling the preregistered report.
- Additional Exploratory analysis (latent class growth analysis) were preregistered.	- These analyses were not reported.	- Convergence issues were encountered and could not be resolved. This may be an indication that the latent growth variables are too closely related. Results were therefore inadmissible.
- Initial preregistered analyses include a set of analyses examining only the interrelationship between ANS, SNS, and translating ability.	- These analyses were calculated but not reported in the main report.	- No theoretically significant differences were observed between these results and the results including math achievement. It was felt the two analyses were therefore redundant.
- Preregistered main analyses were classic cross-lagged panel models (CLPM).	- Main analyses were instead random-intercept CLPMs instead.	- it has been suggested by reviewers that RI-CLPM holds the advantage that stable trait-like factors will be better controlled for.

Appendix 2– Description of Variations performed and Fit Indices for All Models

	Adjusted Score		Unadjusted Score	
	No EDI covariates	EDI covariates	No EDI covariates	EDI covariates
	Mapping Model (I)			
df	185	185	185	185
χ^2	678.459, $p <$ 0.001	551.018, $p <$ 0.001	571.688, $p <$ 0.001	455.520, $p < 0.001$
AIC	39891.913	50320.01	39276.375	49727.663
BIC	41270.557	52452.254	40655.019	51859.908
CFI	0.929	0.949	0.936	0.956
TLI	0.885	0.899	0.896	0.914
RMSEA	0.064	0.055	0.057	0.048
$RMSEA_{Lower}$	0.058	0.049	0.052	0.042
$RMSEA_{Upper}$	0.069	0.06	0.063	0.054
SRMR	0.088	0.049	0.076	0.043
	Mapping Model (II)			
df	182	182	182	182
χ^2	607.001, $p <$ 0.001	489.502, $p <$ 0.001	497.792, $p <$ 0.001	392.992, $p < 0.001$
AIC	39854.71	50285.57	39218.79	49678.83
BIC	41242.22	52426.68	40606.3	51819.94
CFI	0.935	0.954	0.946	0.965
TLI	0.894	0.909	0.911	0.93
RMSEA	0.061	0.052	0.053	0.043
$RMSEA_{Lower}$	0.056	0.047	0.047	0.037
$RMSEA_{Upper}$	0.067	0.058	0.058	0.049
SRMR	0.087	0.048	0.073	0.042
	Refinement Model			
df	185	185	185	185
χ^2	495.146, $p <$ 0.001	390.219, $p <$ 0.001	422.535, $p <$ 0.001	332.589, $p < 0.001$
AIC	39738.86	50182.29	39139.54	49614.42
BIC	41117.5	52314.53	40518.18	51746.67
CFI	0.953	0.969	0.959	0.975
TLI	0.923	0.939	0.934	0.951
RMSEA	0.052	0.042	0.046	0.036
$RMSEA_{Lower}$	0.047	0.037	0.04	0.03
$RMSEA_{Upper}$	0.058	0.048	0.051	0.042
SRMR	0.073	0.039	0.061	0.035
	Time-Dependent / Bi-Directional Model			
df	169	169	169	169
χ^2	445.821, $p <$ 0.001	348.580, $p <$ 0.001	380.338, $p <$ 0.001	293.206, $p < 0.001$
AIC	39721.534	50172.651	39129.339	49607.041

BIC	41171.106	52375.822	40578.91	51810.212
CFI	0.958	0.973	0.964	0.979
TLI	0.925	0.942	0.935	0.955
RMSEA	0.052	0.042	0.045	0.035
$RMSEA_{Lower}$	0.046	0.035	0.039	0.028
$RMSEA_{Upper}$	0.057	0.048	0.051	0.041
SRMR	0.068	0.036	0.059	0.033

Note. Fit Indices of all models under conditions of adjusted vs. unadjusted score and EDI covariates vs. no EDI covariates. While there are differences between model fit between variations, the relationship between the models within variations is very similar.

Note.

Adjusted score and unadjusted score:

Following Lyons et al., (2018), scoring for all tasks were submitted to: A = C – I / (P-1), where A is the adjusted score, C is the total number of correct items, I is the total number of incorrect items, and P is the number of response options (Rowley & Traub, 1977). While this method would adjust for guessing, it is disadvantaged in the sense that the adjusted scores may obscure participants' true performance due to the subtraction (e.g., a child with 3 correct and 1 incorrect will score the same as a child with 13 correct and 11 incorrect). As such, a follow-up analysis was performed for both main and exploratory analyses by using the raw unadjusted scores instead to check for biases that may be introduced by the scoring method. Examination of fit indices and regression coefficients found that the two scoring methods are largely similar. As such, the adjusted score was reported.

Covariate Variation -

All covariates used in the reported models were identical to those used in Lyons et al., (2018). The only exception is the inclusion of EDI covariates in our analyses. We performed separate analyses including and excluding the EDI covariates. We found the EDI covariates frequently, and significantly relate to the main tasks. As such, to better control for the contribution of the domain-general processes measured in the EDI we elected to report the analyses that include the EDI as covariates.

Appendix 3 - Standardized Coefficients of Variables of Interest

	Adjusted Score		Unadjusted Score	
	No EDI covariates	EDI covariates	No EDI covariates	EDI covariates
	Mapping Model (I)			
NC2 ON NC1	0.361***	0.341***	0.359***	0.337***
NC2 ON DC1	0.193***	0.166***	0.252***	0.242***
DC2 ON DC1	0.414***	0.378***	0.426***	0.408***
MC2 ON MC1	0.17***	0.175***	0.122***	0.127***
MC2 ON DC1	0.254***	0.221***	0.294***	0.28***
AV2 ON AV1	0.32***	0.341***	0.239***	0.266***
AV2 ON DC1	0.183***	0.159***	0.178***	0.159***
NC3 ON NC2	0.115**	0.122**	0.152***	0.153***
NC3 ON DC2	0.328***	0.27***	0.229***	0.182***
DC3 ON DC2	0.343***	0.294***	0.258***	0.228***
MC3 ON MC2	**0.075#**	**0.071#**	**0.132****	**0.13****
MC3 ON DC2	0.33***	0.29***	0.184***	0.156**
AV3 ON AV2	0.234***	0.235***	0.264***	0.259***
AV3 ON DC2	**0.191*****	**0.142****	**0.09#**	**0.049#**
AV4 ON AV3	0.483***	0.382***	0.484***	0.361***
AV4 ON DC3	**0.108***	**0.033#**	**0.072#**	**0.027#**
	Mapping Model (II)			
NC2 ON NC1	0.391***	0.369***	0.399***	0.376***
NC2 ON DC1	0.152***	0.126**	0.229***	0.218***
DC2 ON DC1	0.388***	0.348***	0.428***	0.405***
MC2 ON MC1	0.19***	0.193***	0.146***	0.147***
MC2 ON DC1	0.213***	0.179***	0.277***	0.262***
AV2 ON AV1	0.276***	0.311***	0.258***	0.273***
AV2 ON NC1	0.228***	0.206***	0.274***	0.259***
NC3 ON NC2	0.152***	0.155***	0.192***	0.185***
NC3 ON DC2	0.281***	0.238***	0.218***	0.181***
DC3 ON DC2	0.317***	0.277***	0.277***	0.25***
MC3 ON MC2	0.102*	0.094*	0.152***	0.145**
MC3 ON DC2	0.283***	0.257***	0.188***	0.169***
AV3 ON AV2	0.162***	0.167***	0.192***	0.184***
AV3 ON NC2	0.239***	0.219***	0.226***	0.203***
AV3 ON DC1	0.033#	0.009#	-0.021#	-0.035#
AV4 ON AV3	0.432***	0.339***	0.422***	0.325***

AV4 ON NC3	**0.128***	**0.081#**	**0.14****	**0.085#**
AV4 ON DC2	0.054#	-0.017#	0.069#	0.02#
	Refinement Model			
NC2 ON NC1	0.631***	0.599***	0.665***	0.638***
DC2 ON DC1	0.152***	0.153***	0.166***	0.163***
DC2 ON NC1	0.432***	0.401***	0.412***	0.392***
MC2 ON MC1	0.124***	0.127***	0.136***	0.137***
MC2 ON NC1	0.456***	0.437***	0.433***	0.423***
AV2 ON AV1	0.294***	0.315***	0.228***	0.239***
AV2 ON NC1	0.346***	0.322***	0.423***	0.403***
NC3 ON NC2	0.466***	0.428***	0.475***	0.428***
DC3 ON DC2	0.025#	0.026#	0.04#	0.045#
DC3 ON NC2	0.391***	0.349***	0.356***	0.316***
MC3 ON MC2	0.123**	0.125**	0.121**	0.126**
MC3 ON NC2	0.331***	0.287***	0.336***	0.292***
AV3 ON AV2	0.159***	0.162***	0.167***	0.161***
AV3 ON NC2	0.368***	0.328***	0.352***	0.305***
AV4 ON AV3	0.472***	0.372***	0.452***	0.349***
AV4 ON NC3	**0.148****	**0.074#**	**0.168*****	**0.096***
	Time Dependent / Bidirectional Model			
NC2 ON NC1	0.535***	0.512***	0.572***	0.545***
NC2 ON DC1	-0.004#	-0.022#	0.078#	0.072#
NC2 ON MC1	**0.159*****	**0.155*****	**0.081#**	**0.079#**
DC2 ON DC1	0.161**	0.141**	0.211***	0.196***
DC2 ON NC1	0.382***	0.361***	0.349***	0.331***
DC2 ON MC1	0.083#	0.077#	0.066#	0.063#
MC2 ON MC1	0.214***	0.213***	0.174**	0.174**
MC2 ON NC1	0.396***	0.388***	0.367***	0.359***
MC2 ON DC1	0.011#	-0.011#	0.081#	0.066#
AV2 ON AV1	0.307***	0.323***	0.269***	0.273***
AV2 ON NC1	0.346***	0.325***	0.431***	0.42***
AV2 ON MC1	0.011#	0.006#	0.047#	0.043#
AV2 ON DC1	-0.005#	-0.014#	-0.07#	-0.085#
NC3 ON NC2	0.313***	0.308***	0.425***	0.393***
NC3 ON DC2	**0.125***	**0.095#**	**0.013#**	**-0.014#**
NC3 ON MC2	0.109#	0.093#	0.068#	0.069#
DC3 ON DC2	0.121#	0.096#	0.028#	0.016#
DC3 ON NC2	0.29***	0.27***	0.36***	0.328***

DC3 ON MC2	0.067#	0.056#	0.021#	0.02#
MC3 ON MC2	0.171**	0.16*	0.166*	0.166*
MC3 ON NC2	0.214***	0.191***	0.34***	0.301***
MC3 ON DC2	**0.134***	**0.114#**	**-0.04#**	**-0.05#**
AV3 ON AV2	0.161***	0.165***	0.207***	0.197***
AV3 ON NC2	0.316***	0.294***	0.414***	0.378***
AV3 ON MC2	**0.151***	**0.146***	**0.075#**	**0.083#**
AV3 ON DC2	**-0.093#**	**-0.112#**	**-0.184****	**-0.2****
AV3 ON DC1	0.026#	0.007#	-0.011#	-0.025#
AV4 ON AV3	0.486***	0.401***	0.472***	0.372***
AV4 ON NC3	0.135*	0.113*	0.159**	0.109*
AV4 ON MC3	**-0.116#**	**-0.119***	**-0.072#**	**-0.066#**
AV4 ON DC3	0.078#	0.036#	0.023#	0.014#
AV4 ON DC2	**0.094***	**0.02#**	**0.102***	**0.045#**

Note. Standardized coefficients of auto-regressive and model specific regressions. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, $^{\#}$ $p > 0.05$. Bolded rows indicate substantial disagreement in regression significance between the variations.

Addendum. Of the 79 regression coefficients of interest, 12 had substantial disagreement in the regression coefficient significance. Most disagreements occur when both ANS and SNS are both allowed to affect later performance (Time-Dependent / Bi-Directional Model), while the more restrictive models tended to have a high degree of agreement. This may be an indication of overfitting of the Time-Dependent / Bi-Directional Model, and that the estimated coefficients may be capitalizing on chance associations between variables. As such, one may be recommended to take the conclusions reached by the Time-Dependent / Bi-Directional Model with some of skepticism.

Appendix 4 - Additional Models Considered

After completing the preregistered models in the main analyses, we considered additional variations of the mapping and refinement accounts to explore uncertainties of extant research on the intercorrelations between the four main variables. Please note that all models described below were done as post-hoc exploratory analyses.

Mapping Account - Translating Ability Orthogonal to the ANS

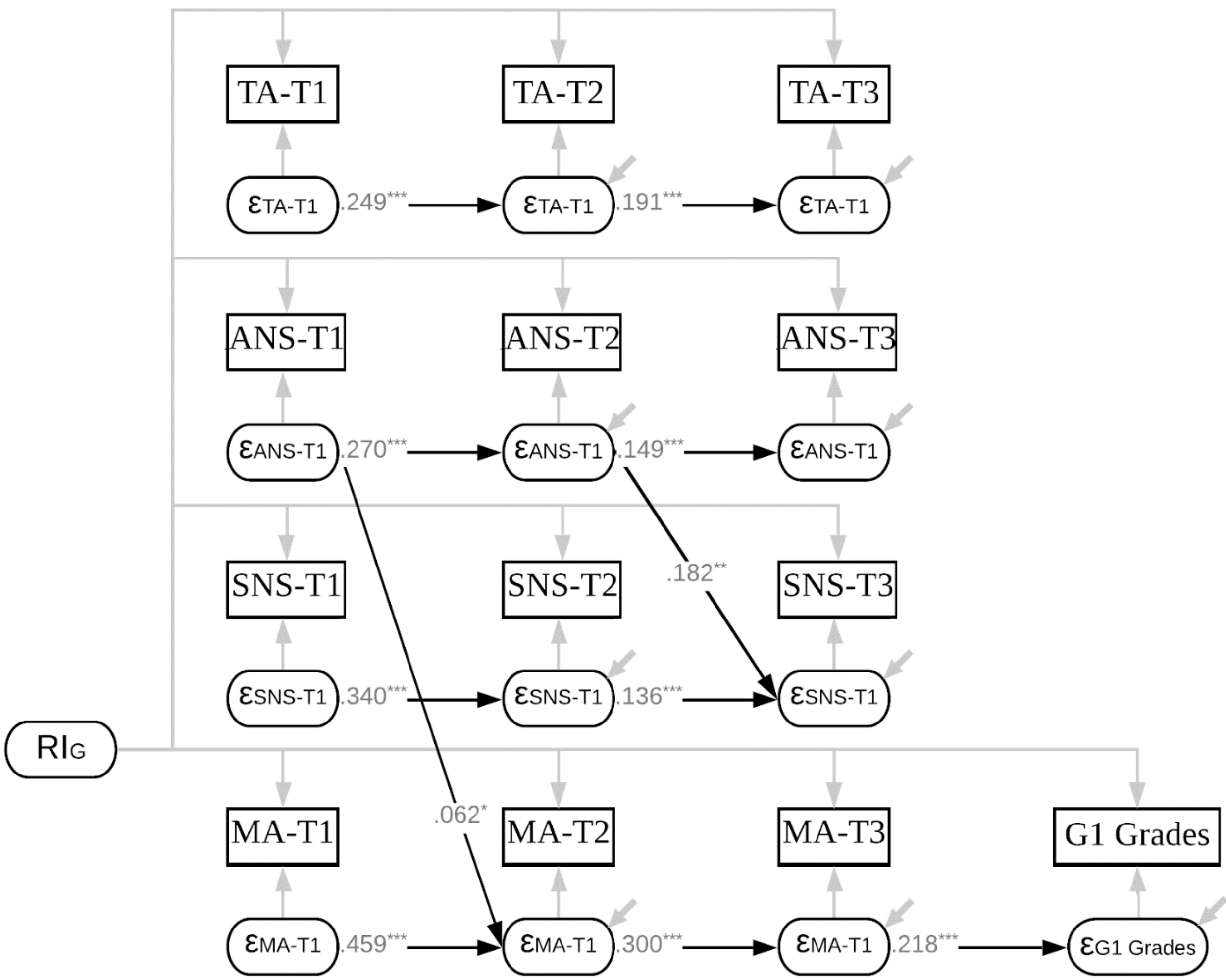
It has been argued that children with stronger ANS abilities would also have a stronger mapping between the ANS and symbolic numerals, presumably because students with stronger ANS abilities will have an easier time distinguishing between representations of neighboring symbolic numerals (Geary, 2013; Mussolin et al., 2012). However, evidence is somewhat mixed with regards to whether the ANS may underlie translating ability. While some studies have found evidence to suggest there is a relationship between ANS acuity and translating ability (Brankaer et al., 2014; Jang & Cho, 2018; Pinheiro-Chagas et al., 2014; Wong et al., 2016), several studies have found no relationship between translating ability and ANS acuity (Gimbert et al., 2019; Libertus et al., 2014).

To reflect on this uncertainty, we have compared Mapping Model (I) and (II) against modified versions of the models whereby translating ability is not predicted by earlier ANS. The modified Mapping Models were compared with the original model. Results indicate model fit was adequate, for both modified Mapping Model (I), $\chi^2_{(186)}$ = 576.206, p < 0.001, AIC = 50364.277, BIC = 52487.655, CFI = 0.942, TLI = 0.887, RMSEA = 0.058 [90% CI: 0.053, 0.063], SRMR = 0.055, and modified Mapping Model (II) $\chi^2_{(184)}$ = 533.337, p < 0.001, AIC = 50325.408, BIC = 52457.652, CFI = 0.948, TLI = 0.897, RMSEA = 0.055 [90% CI: 0.050, 0.061], SRMR = 0.052.

Results indicate that the removal of the link between ANS and translating ability had inferior fit for both Mapping Model (I), ΔAIC = 44.267, ΔBIC = 35.401, $\chi^2(2) = 48.267$, $p < 0.001$, and Mapping Model (II), ΔAIC = 30.835, ΔBIC = 30.969, $\chi^2(2) = 43.835$, $p < 0.001$.

Figure A4-1

Mapping Account - Translating Ability Orthogonal to the ANS



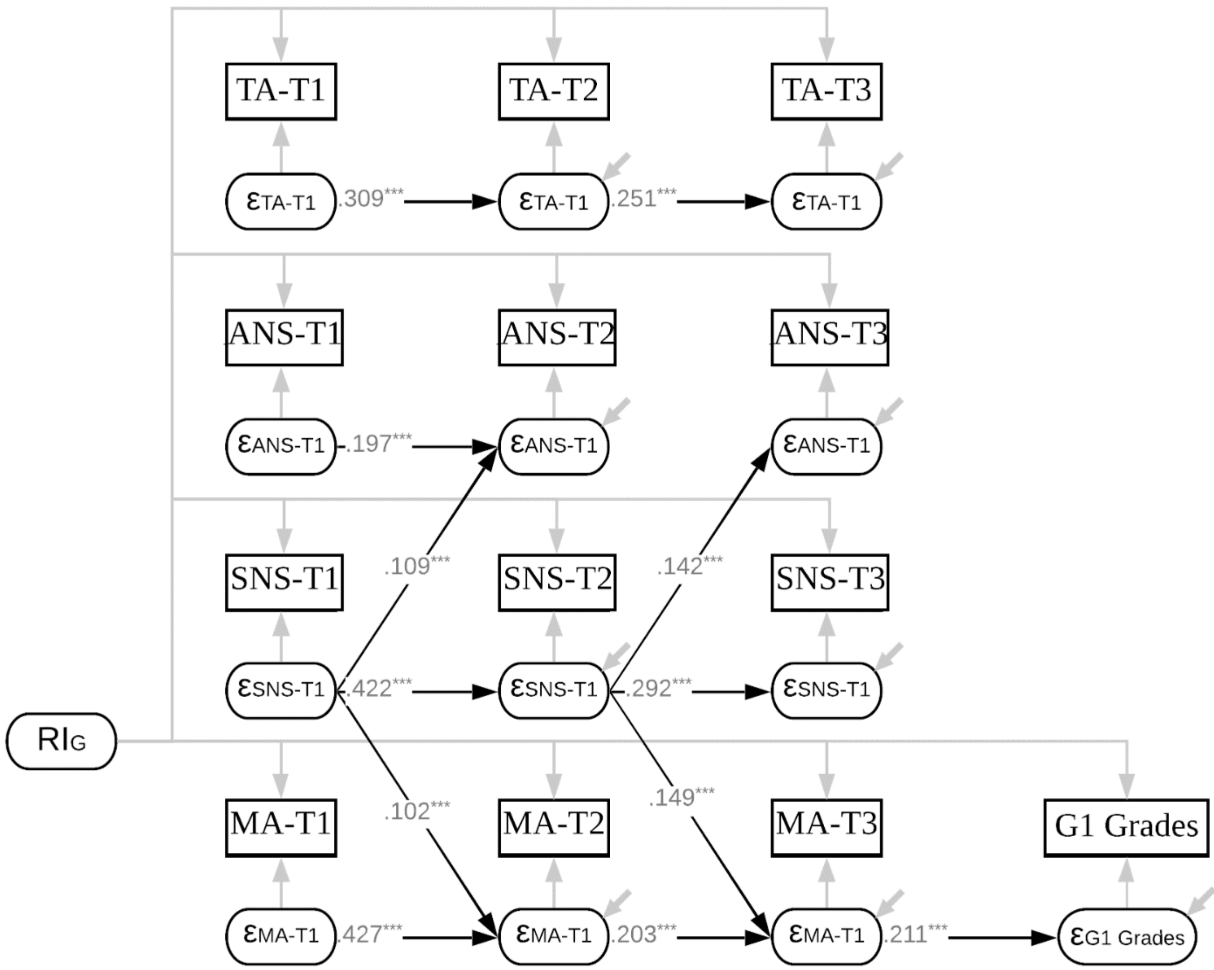
Note. TA refers to translating ability and MA refers to math achievement. T1-3 refers to time-points of data collection. Grey lines are specifications for the random intercepts and structured residuals. Black lines are the autoregressive and cross-lagged coefficients of interest. Statistically insignificant path coefficients are not shown for clarity. Grey lines are specifications for the random intercepts and structured residuals and are not numerated for clarity.

Refinement Account - Translating Ability Orthogonal to the SNS

We have hypothesized that growth in translating ability would largely be dependent on familiarity and practice with translating between ANS and SNS representations. As such, in early childhood, increased accuracy in translation between ANS and SNS would be tightly linked with the learning of symbolic numerals because both are underwritten by the same pedagogical tasks that supports the initial learning of symbolic numerals (e.g., counting objects; Carey, 2004; Mix, 2008; Mussolin et al., 2014, 2016). However, as there are few extant research papers that has examined the relationship between translating ability and SNS acuity, we have considered the refinement model whereby later translating ability was not predicted by earlier SNS. Results indicate that model fit was adequate, $\chi^2_{(186)} = 511.590$, $p < 0.001$, AIC = 50299.660, BIC = 52423.039, CFI = 0.952, TLI = 0.905, RMSEA = 0.053 [90% CI: 0.048, 0.059], SRMR = 0.050. However, the modified refinement model had inferior fit to the original model, ΔAIC = 117.37, ΔBIC = 108.505, $\chi^2(2) = 121.371$, $p < 0.001$.

Figure A4-2

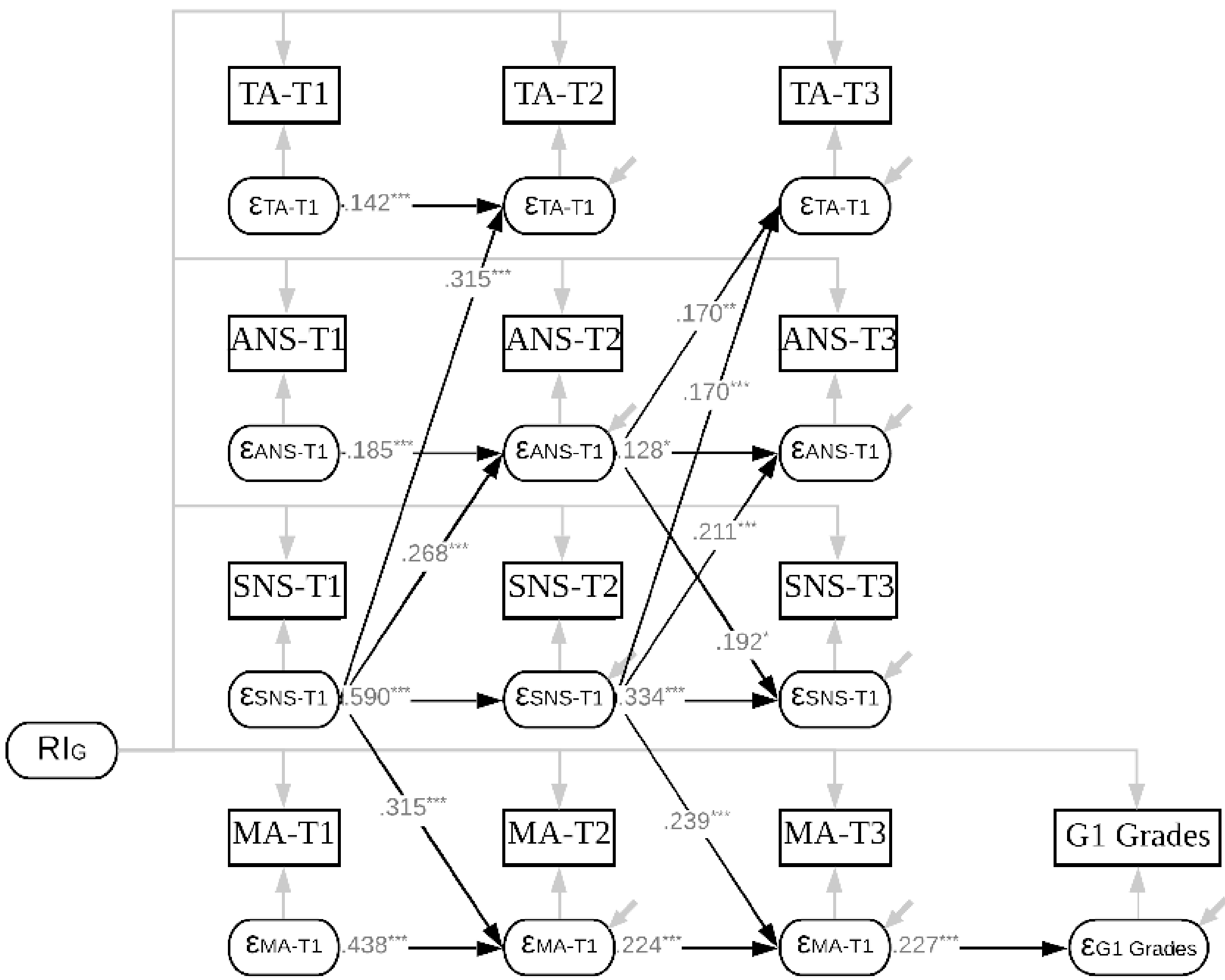
Mapping Account - Translating Ability Orthogonal to the ANS



Note. TA refers to translating ability and MA refers to math achievement. T1-3 refers to time-points of data collection. Grey lines are specifications for the random intercepts and structured residuals. Black lines are the autoregressive and cross-lagged coefficients of interest. Statistically insignificant path coefficients are not shown for clarity. Grey lines are specifications for the random intercepts and structured residuals and are not numerated for clarity.

Figure A4-3

Bidirectional / Transient Model - Translating Ability Orthogonal to ANS and SNS



Note. TA refers to translating ability and MA refers to math achievement. T1-3 refers to time-points of data collection. Grey lines are specifications for the random intercepts and structured residuals. Black lines are the autoregressive and cross-lagged coefficients of interest. Statistically insignificant path coefficients are not shown for clarity. Grey lines are specifications for the random intercepts and structured residuals and are not numerated for clarity.

Appendix 5 - Additional Statistical Considerations

Classroom and School Level Variations in the TIMSS and PISA

One major difference between the sampling methodologies of the TIMSS and PISA datasets is the primary sampling unit. While the primary sampling unit of the TIMSS dataset is the classroom while the primary sampling unit for the PISA is the school (OECD, 2014a). As such, interpretations of level-2 variations for the TIMSS and PISA models would be different. Specifically, variations at level-2 can mostly be attributed to the classroom for the TIMSS and the school for the PISA.

The omitted levels for all three ML-SEMs would introduce imprecision in the estimated variations that exists at different levels. Simulation studies have shown the variance attributed to an omitted level is divided between the flanking levels (Van den Noortgate et al., 2005). As such, the omission of the school level for the TIMSS would inflate variance attributed to the classroom and country levels, while the omission of the classroom level for the PISA would inflate variance attributed to the individual and school levels.

In both cases the omission of a level inflates variations attributed to level-2 in the ML-SEM modelled in the current study. While the statistically significant average contextual effect at level-2 in all three ML-SEMs suggests that the math anxiety level of students' immediate educational environment (i.e., the school and the classroom) have an addictive detrimental effect on students' math achievement, the specific level of the contextual effect (i.e., whether the effect is at the classroom level or school level or both) cannot be determined from the data. Whether the contextual effect exists on the classroom- or school-levels or both is a subject for future study.

Sampling Weights

Both the TIMSS and the PISA employ a stratified two-stage random sample design (Martin et al., 2016; OECD, 2014a). The usage of sampling weights is required to adjust for the non-independency introduced by the sampling design (Carle, 2009). For the TIMSS, sample weights for the student level were calculated as the product of WGTFAC3 and WGTADJ03, and sample weights for the classroom level were calculated as the product of WGTFAC01, WEGADJ01, WGTFAC02, and WTGADJ02 (Rutkowski et al., 2010). For the PISA, sample weights for the student level was computed as the ratio between the final student weight and the final school weight (i.e., W_FSTUWT/ W_FSCHWT) and sample weights for the school level was the final school weight (W_FSCHWT; Carle, 2009; Nagengast & Marsh, 2011; Stapleton, 2006).

Measurement Model for the Latent Variables in TIMSS Grade 4

Multiple covariates of math anxiety and math achievement in the final model contained multiple items and were operationalized as latent variables. Specifically, latent variables at the student level were student attitudes towards the math teacher, student attitudes towards the school, student attitudes towards mathematics, current extracurricular tutoring/lessons, home math activities during preschool, and parental attitudes towards math and science. Latent variables at the teacher level were teacher confidence in teaching math and teacher satisfaction with work (see appendix 6 & 7 for item descriptions). The factor structure of these latent variables was examined with a measurement model CFA, one for each level. Results indicate that latent variables at both individual level, $\chi^2_{(650)} = 56529.33$, $p < .001$, CFI = .990, TLI = .989, RMSEA = .015, SRMR = .045, and classroom level, $\chi^2_{(89)} = 592.77$, $p < .001$, CFI = .977, TLI = .973, RMSEA = .004, SRMR = .028, were well fitting.

Model Fit

Model fit were reported where available. However, several models were saturated or contained random slopes. In both cases, conventional model fit indices were not available.

Appendix 6 – Variable Description

Table A6-1. Item description of variables.

Variable	Item Description
Individual Level	
Math Anxiety	TIMSS Grade 4: *ASBM03E – Mathematics makes me nervous TIMSS Grade 8: *BSBM19E – Mathematics makes me nervous PISA: * ST42Q01 – I often worry that it will be difficult for me in mathematics classes * ST42Q03 – I get very tense when I have to do mathematics homework * ST42Q05 – I get very nervous doing mathematics problems * ST42Q08 – I feel helpless when doing a mathematics problem
Student Gender	ASBG01 - Are you a girl or a boy? 1) Girl, 2) Boy
Student Attitudes towards the Math Teacher	*ASBM02A – I know what my teacher expects me to do *ASBM02B – My teacher is easy to understand *ASBM02C – I am interested in what my teacher says *ASBM02D – My teacher gives me interesting things to do *ASBM02E – My teacher has clear answers to my questions *ASBM02F – My teacher is good at explaining mathematics *ASBM02G – My teacher lets me show what I have learned *ASBM02H – My teacher does a variety of things to help us learn *ASBM02I – My teacher tells me how to do better when I make a mistake

	*ASBM02J – My teacher listens to what I have to say
Student Attitudes towards the School	*ASBG11A – I like being in school *ASBG11B – I feel safe when I am at school *ASBG11C – I feel like I belong at this school *ASBG11D – I like to see my classmates at school *ASBG11E – Teachers at my school are fair to me *ASBG11F – I am proud to go to this school *ASBG11G – I learn a lot in school
Student Attitudes towards Math	
Years of Preschool Education	ASBH05B - Approximately, how long was your child in these programs [referring to pre-primary education] altogether? 1) Did not attend, 2) Less than 1 year, 3) 1 year, 4) 2 years, 5) 3 years, 6) 4 years or more
Home Math Activities during Preschool	Before your child began primary/elementary school, how often did you or someone else in your home do the following activities with him or her? *ASBH02J – Say counting rhymes or sing counting songs *ASBH02K – Play with number toys (e.g., blocks with numbers) *ASBH02L – Count different things *ASBH02M – Play games involving shapes (e.g., shape sorting toys, puzzles) *ASBH02N – Play with building blocks or construction toys *ASBH02O – Play board or card games *ASBH02P – Write numbers 1) Often, 2) Sometimes, 3) Never or almost never

Current Extracurricular Tutoring/Lessons	ASBH10BA – For how many of the last 12 months has your child attended extra lessons or tutoring? 1) Did not attend, 2) Less than 4 months, 3) 4-8 months, 4) More than 8 months
Parental Involvement in Homework	How often do you or someone else in your home do the following things? *ASBH09BB –Help your child with homework *ASBH09BC –Review your child's homework to make sure it is correct 1) Every day, 2) 3 or 4 times a week, 3) 1 or 2 times a week, 4) Less than once a week, 5) Never or almost never
Parental Attitudes towards Math	How much do you agree with these statements about mathematics and science? ASBH16A – Most occupations need skills in math, science, or technology ASBH16D – My child needs mathematics to get ahead in the world ASBH16G – Mathematics is applicable to real life ASBH16H – Engineering is necessary to design things that are safe and useful
Parent's Highest Education Level	ASDHEDUP – Highest level of education of either parent 1) Finished some primary or lower secondary or did not go to school, 2) Finished lower secondary, 3) Finished upper secondary, 4) Finished post-secondary education, 5) Finished university or higher
Parent's Occupation	*ASDHOCCP – Highest level of occupation of either parent

	1) Professional, 2) Small Business Owner, 3) Clerical, 4) Skilled Worker, 5) General Laborer, 6) Never worked outside home, 7) Not Applicable [7 recoded as missing]
Home Socioeconomic Status	ASBG04 – Number of books in the home: 1) 0-10, 2) 11-25, 3) 26-100, 4) 101-200, 5) More than 200
Classroom Level	
Weekly Time Spent on Math	ATBM01 – In a typical week, how much time do you spend teaching mathematics to the students in this class? (minutes) Minutes in Integer
Frequency of Mix-Ability Grouping	ATBM03H – In teaching mathematics to this class, how often do you ask students to do the following? Work in mixed ability groups 1) Every or almost every lesson, 2) About half the lessons, 3) Some lessons, 4) Never
Frequency of Same-Ability Grouping	ATBM03I – In teaching mathematics to this class, how often do you ask students to do the following? Work in same ability groups 1) Every or almost every lesson, 2) About half the lessons, 3) Some lessons, 4) Never
Homework Frequency	ATBM07B – When you assign mathematics homework to the students in this class, about how many minutes do you usually assign? (Consider the time it would take an average student in your class.) 1) 15 minutes or less, 2) 16-30 minutes, 3) 31-60 minutes, 4) more than 60 minutes
Teacher Gender	ATBG02 – Are you female or male?

	1) Female, 2) Male
Teacher Satisfaction with Work	How often do you feel the following way about being a teacher? *ATBG10A – I am content with my profession as a teacher *ATBG10B –I am satisfied with being a teacher at this school *ATBG10C – I find my work full of meaning and purpose *ATBG10D – I am enthusiastic about my job *ATBG10E – My work inspires me *ATBG10F – I am proud of the work I do *ATBG10G – I am going to continue teaching for as long as I can 1) Very often, 2) Often, 3) Sometimes, 4) Never or almost never
Teacher Confidence in Teaching Math	In teaching mathematics to this class, how would you characterize your confidence in doing the following? *ATBM02A – Inspiring students to learn mathematics *ATBM02B – Showing students a variety of problem solving strategies *ATBM02C – Providing challenging tasks for the highest achieving students *ATBM02D – Adapting my teaching to engage students' interest *ATBM02E – Helping students appreciate the value of learning mathematics *ATBM02F – Assessing student comprehension of mathematics *ATBM02G – Improving the understanding of struggling students *ATBM02H – Making mathematics relevant to students *ATBM02I – Developing students' higher-order thinking skills 1) Very high, 2) High, 3) Medium, 4) Low

Teacher Years of Teaching Experience	ATBG01 – By the end of this school year, how many years will you have been teaching altogether? Integer response.
Teacher Years of Formal Education	ATBG04 – What is the highest level of formal education you have completed? 1) Did not complete <Upper secondary education—ISCED Level 3>, 2) <Upper secondary education— ISCED Level 3>, 3) <Post-secondary, non-tertiary education—ISCED Level 4>, 4) <Short-cycle tertiary education—ISCED Level 5>, 5) <Bachelor's or equivalent level—ISCED Level 6>, 6) <Master's or equivalent level—ISCED Level 7>, 7) <Doctor or equivalent level—ISCED Level 8>
Teacher Major	ATDM05 –Teachers Majored in Education and Mathematics [Derived Variable] 1) Major in primary education and major (or specialization) in mathematics, 2) Major in primary education but no major (or specialization) in mathematics, 3) Major in mathematics but no major in primary education, 4) All other majors 5) No formal education beyond upper-secondary [Recoded into binary variable, 1 being major in education or mathematics, 0 being did not major in education or mathematics]
Country-Level	
Human Development Index	The human development index is a composite score of a country's development measured across three broad dimensions: long and

	healthy life, knowledge, and a decent standard of living. The health dimension is assessed by the life expectancy a typical citizen at birth. The education dimension is measured by the average years of schooling for adults above the age of 25, and the expected years of schooling for children at the first year of schooling. The standard living dimension is measured by the gross national income. The three dimensions are combined into a composite score by using geometric mean (United Nations Development Program, 2015).
Individualism-Collectivism Index	Individual-Collectivism is a bipolar variable that refers the degree to which ties between individuals are loose (individualistic) or are tight (collectivist). An example question is, “Try to think of those factors that would be important to you in an ideal job; disregard the extent to which they are contained in your present job. How important is it to you to have sufficient time for your personal or family life?” (from 1 “of utmost importance to me” to 5 “of very little or no importance”). High values of individualism-collectivism index indicate higher individualism, and lower values indicate higher collectivism. Example taken from Hofstede et al. (2010).
Power Distance Index	Power distance refers to the degree to which the less powerful members of an organization expect and accept that power is distributed unequally An example question is, “How frequently, in your experience, does the following problem occur: employees being afraid to express disagreement with their managers?” (from “very frequently” to “very seldom”). High values of power distance index indicate less acceptance of inequalities with a country. Example taken from Hofstede et al. (2010).
Uncertainty Avoidance Index	Uncertainty avoidance refers to the degree to which the members of a culture feels threatened by ambiguous or unknown situations. An example question is, “Company rules should not be broken—even when the employee thinks it is in the company’s best interest” (from

	1 “I always feel this way” to 5 “I never feel this way.”) High values of uncertainty avoidance indicate less acceptance of uncertainties. Example taken from Hofstede et al. (2010).
Masculinity Index	Masculinity is a bipolar variable that refers to degree to which emotional roles are clearly distinct. Multiple masculine and feminine values are given in Hofstede et al. (2010). The following statements are most pertinent to the classroom. Feminine values include: “average student is the norm; praise for weak students,” “jealousy of those wo try to excel,” “failing in school is a minor incident, “students underrate their own performance: ego-effacement,” and “friendliness in teachers is appreciated.” Masculine values include: “best student is the norm; praise for excellent students,” “competition in class; trying to excel,” “failing in school is a disaster,” “students overrate their own performance: ego-boosting,” and “brilliance in teachers is admired.” Higher values of masculinity indicate higher degree of endorsement of masculine values, while lower values of masculinity indicate higher degree of endorsement of feminine values.
Long-Term Orientation Index	Long-term orientation is a bipolar variable that refers to the degree to which members of a culture are oriented towards future or to the past and present. Particularly, long-term oriented cultures endorse perseverance and thrift, while short-term oriented cultures endorse respect for tradition and the preservation of “face.” Higher values of long-term orientation indicate a higher level of orientating towards the future, while lower values of long-term orientation indicate a higher degree of orienting towards the past and present.

* reverse coded. All items correspond to a 4-point Likert scale with 1 being agree a lot and 4 being disagree a lot unless otherwise specified. All items are found in the TIMSS Grade 4 unless otherwise specified. Binary categorical variables are recoded to be 0 and 1.

Appendix 7 - Substantive Basis for the Included Variables for TIMSS Grade 4

Individual Level Variables

Student and Teacher Gender

Multiple studies have found that adult females generally experience higher levels of math anxiety (Dew & Galassi, 1983; Ferguson et al., 2015; Hembree, 1990; Miller & Bichsel, 2004; Woodard, 2004). However, evidence is more mixed in studies with children. While some studies have reported no gender difference in math anxiety (Gierl & Bisanz, 1995; Harari et al., 2013; Ramirez et al., 2013) other studies have found that, similar to adults, girls experience higher levels of anxiety (Griggs et al., 2013; Satake & Amato, 1995; Yüksel-Şahin, 2008). Interestingly, there may be an interaction between a child's gender and the teacher's gender. For instance, Beilock and colleagues have found gender specific effects of math anxiety whereby it is female students of math anxious female teachers to be more likely to be behind in math (Beilock et al., 2010). However, a larger replication study have seem to suggest that both genders may experience lower math performance (Schaeffer et al., 2020).

Student Attitudes

Extant research has revealed that student attitudes are strong predictors of math anxiety. For instance, students' own beliefs regarding their own math ability has been shown to be inversely associated with math anxiety (e.g., Ahmed et al., 2012; Lee, 2009). Similarly, students' positive beliefs regarding the teacher efficacy and the school learning environment atmosphere have been shown to be negatively associated with math anxiety (e.g., Fast et al., 2010; Radišić et al., 2015). A more positive opinion regarding the teacher and learning environment may be associated with reduced math anxiety because stronger support from teachers and peers may help reduce negative expectations that could induce math anxiety.

Years in Pre-primary Education and Home Math Activities During Preschool

Not much is known regarding whether preschool attendance would have a positive or negative effect on math anxiety. Multiple studies have found that preschool teachers may exhibit math anxiety (e.g., Aslan, 2013) and teachers' anxiety may be transmitted to students (Beilock et al., 2010). Indeed, there is some support for the notion that children at that age already experience math anxiety and that it affects their math achievement (Lu et al., 2019; Stipek & Ryan, 1997). However, few research studies have examined whether the attendance of pre-primary education would yield a net positive or negative effect on math anxiety when compared with not attending pre-primary education. On the one hand, earlier exposure to possible instigators of math anxiety would have one predict an association between pre-primary education and math anxiety. Specifically, one would predict that more years of pre-primary education would lead to more math anxiety. On the other hand, pre-primary education may endow a protective effect by better equipping students for learning of formal mathematical concepts in the first grade, which suggests the more pre-primary education, the less math anxiety students will feel.

Similarly, home math activities during preschool may be related to math anxiety as it may endow a protective effect by preparing students for learning mathematics in the first grade. Interestingly, evidence is mixed, some studies have found home activities seem to reduce math anxiety (Berkowitz et al., 2015), with other studies finding that home activities increases math anxiety (Jameson, 2014)**.** However, extant studies tend to measure home math activities concurrent with the measurement of math anxiety. To the best of our ability, we have not found studies that examine how home math activities prior to formal schooling would affect later math anxiety.

Parental Involvement with Homework and Current Extracurricular Tutoring/Lesson

Surprisingly, extant research examining homework have generally found a negative effect of the amount of homework on math anxiety. For instance, a higher proportion of total homework time spent doing math homework is associated with higher math anxiety (Cheema & Sheridan, 2015; Radišić et al., 2015). It is uncertain whether this may a reflect of children experiencing more difficulty with doing more math homework or have a higher quantity of math homework. Further, there is evidence to suggest that parents with high math anxiety who also help their children with homework seem to increase their children’s math anxiety (Maloney et al., 2015). Taken together, math homework can be an indicator of student who are already struggling or can be considered a vehicle through which parental math anxiety may be transmitted to students.

In contrast, few research studies have examined the effects of extra-curricular tutoring on math anxiety. Existing studies have designed and implemented intensive and specialized math tutoring session to general positive effect (Supekar et al., 2015). However, the effects of normal extracurricular tutoring have not been examined. The effects of extra-curricular tutoring may be complicated due to its relationship with student SES – that is, those students with higher SES may have a higher likelihood of receiving extra-curricular tutoring. However, similar with homework, extra-curricular tutoring may be an indicator of students who are already struggling. Taken together, one may expect a negative relationship between extra-curricular tutoring and math anxiety, but a positive relationship once SES is statistically accounted.

Parental Attitudes towards Math and Science

Parents serve as role models and influences children's perception and attitudes towards math. For instance, children of more math anxious teachers tend to endorse gender stereotypes in mathematics (Beilock et al., 2010), and parental support and encouragement is positively related with student attitudes towards achievement in school (Onslow, 1993; Parsons et al., 1982). Indeed, there is evidence to suggest that parental attitudes regarding the importance of mathematics seem to be associated with lowered math anxiety (Soni & Kumari, 2017). Taken together, we should expect positive parental attitudes towards math and science to be related with lowered math anxiety.

Parents' Highest Education, Occupation and Home Socioeconomic Status

The evidence for the effects of socioeconomic status on math anxiety is somewhat mixed with some research suggesting a negative relationship (Baya'a, 1990; Cheema & Galluzzo, 2013; Geyik, 2015) and other studies finding no relationship (Radišić et al., 2015). Further, research has found that parental education is inversely related to math anxiety (Ahmed, 2018) and parental education can be considered a stable indicator of home SES (Sirin, 2005). Taken together, evidence suggest that SES may be inversely related with student math anxiety, but some studies failed to find such a relation.

Classroom Level Variables

Frequency of Mixed-Ability and Same-Ability Grouping in the Classroom

To the best of our ability, we could not find extant studies that examine whether mixed- or same-ability groupings in the classroom would affect student math anxiety. However, research that examine the relationship between ability grouping and student attitudes seem to suggest that mixed ability grouping generally lead to more student positive outlook on math. In contrast, same-ability grouping seems to lead to more negative outlook on math (Boaler et al., 2000; Catsambis et al., 2001). This suggests the ability of one's peers may affect student attitudes towards mathematics. In light of this, it is quite plausible to expect ability grouping may affect math anxiety. For instance, one may expect that since lower ability children generally experience more math anxiety, same ability grouping may exacerbate the experience of math anxiety due to children being exposed to peers who are similarly anxious about math.

Weekly Time Spent on Math and Homework Frequency

See Parental Involvement with Homework.

Teacher Gender

See student gender.

Teacher Confidence in Teaching Math

A multitude of extant studies have examined how teachers may influence student math anxiety. For example, a number of studies have shown that teachers often experience math anxiety and that students with teachers who experience math anxiety may have lower math achievement (Novak & Tassell, 2017; Ramirez et al., 2018).

While the TIMSS Grade 4 do not include measurement of teacher anxiety, teacher confidence in teaching math is included. As previous research has indicated that teacher math anxiety is a strong negative predictor of teacher confidence in teaching math (Bursal & Paznokas, 2006; Geist, 2015), it is reasonable to hypothesize that teacher confidence may also be related with student math anxiety and achievement.

Teacher Years of Teaching Experience, Teacher Years of Formal Education, Teacher Major, and Teacher Satisfaction with Work

One important predictor of student math anxiety is classroom atmosphere (Radišić et al., 2015). For instance, it has been suggested that certain pedagogical practice, such as, assuming many parts of the mathematical procedures to be simple and self-explanatory, using unique vocabulary without explanation, and overusing rote drills, may lower student confidence and increase student math anxiety (Cornell, 1999). Indeed, negative experience with the teacher seems to be prominent way through which students acquire math anxiety (Jackson & Leffingwell, 1999).

While pedagogical practice differs between teachers, years of teaching experience, formal education, major, and satisfaction with work may serve as potential indicators for a teacher's ability to inspire a positive atmosphere in the classroom.

Country-Level Variables

Human Development Index and Cultural Dimensions

It has been previously observed that different countries vary widely in terms of the average student math anxiety, and it has bee postulated that cultural context may play a role in this between-country difference (Foley et al., 2017). Specifically, it has been proposed that between country differences in the parents' role in a child's mathematics education, and differences in competitiveness in the education environment may contribute to the observed differences (Foley et al., 2017). To control for between country differences in socio-economic status, the UN human development index is used. To control for between country differences in the behaviors of parents and teachers, Hofstede et al. (2010)'s cultural dimensions are used.

References

Ahmed, W. (2018). Developmental trajectories of math anxiety during adolescence: Associations with STEM career choice. Journal of Adolescence, 67, 158–166.

Ahmed, W., Minnaert, A., Kuyper, H., & van der Werf, G. (2012). Reciprocal relationships between math self-concept and math anxiety. Learning and Individual Differences, 22(3), 385–389.

Aslan, D. (2013). A Comparison of Pre- and In-Service Preschool Teachers' Mathematical Anxiety and Beliefs about Mathematics for Young Children. Academic Research International, 4(2), 225.

Bayaʼa, N. F. (1990). Mathematics anxiety, mathematics achievement, gender, and socio‐economic status among Arab secondary students in Israel. International Journal of Mathematical Education in Science and Technology, 21(2), 319–324.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. Proceedings of the National Academy of Sciences, 107(5), 1860–1863.

Berkowitz, T., Schaeffer, M. W., Maloney, E. A., Peterson, L., Gregor, C., Levine, S. C., & Beilock, S. L. (2015). Math at home adds up to achievement in school. Science, 350(6257), 196–198.

Boaler, J., Wiliam, D., & Brown, M. (2000). Students' experiences of ability grouping-disaffection, polarisation and the construction of failure. British Educational Research Journal, 26(5), 631–648.

Bursal, M., & Paznokas, L. (2006). Mathematics anxiety and preservice elementary teachers' confidence to teach mathematics and science. School Science and Mathematics, 106(4), 173–180.

Catsambis, S., Mulkey, L., & Crain, R. (2001). For better or for worse? A nationwide study of the social psychological effects of gender and ability grouping in mathematics. Social Psychology of Education, 5(1), 83–115.

Cheema, J. R., & Galluzzo, G. (2013). Analyzing the gender gap in math achievement: Evidence from a large-scale US sample. Research in Education, 90(1), 98–112.

Cheema, J. R., & Sheridan, K. (2015). Time spent on homework, mathematics anxiety and mathematics achievement: Evidence from a US sample. Issues in Educational Research, 25(3), 246.

Cornell, C. (1999). I hate math! I couldn't learn it, and I can't teach it! Childhood Education, 75(4), 225.

Dew, K. H., & Galassi, J. P. (1983). Mathematics anxiety: Some basic issues. Journal of Counseling Psychology, 30(3), 443.

Fast, L. A., Lewis, J. L., Bryant, M. J., Bocian, K. A., Cardullo, R. A., Rettig, M., & Hammond, K. A. (2010). Does math self-efficacy mediate the effect of the perceived classroom environment on standardized math test performance? Journal of Educational Psychology, 102(3), 729.

Ferguson, A. M., Maloney, E. A., Fugelsang, J., & Risko, E. F. (2015). On the relation between math and spatial ability: The case of math anxiety. Learning and Individual Differences, 39, 1–12.

Foley, A. E., Herts, J. B., Borgonovi, F., Guerriero, S., Levine, S. C., & Beilock, S. L. (2017). The math anxiety-performance link: A global phenomenon. Current Directions in Psychological Science, 26(1), 52–58.

Geist, E. (2015). Math anxiety and the “math gap”: How attitudes toward mathematics disadvantages students as early as preschool. Education, 135(3), 328–336.

Geyik, S. K. (2015). The effects of parents’ socio economic status on mathematics anxiety among social sciences students in turkey. International Journal of Education and Research, 3(1), 311–324.

Gierl, M. J., & Bisanz, J. (1995). Anxieties and attitudes related to mathematics in grades 3 and 6. The Journal of Experimental Education, 63(2), 139–158.

Griggs, M. S., Rimm-Kaufman, S. E., Merritt, E. G., & Patton, C. L. (2013). The Responsive Classroom approach and fifth grade students’ math and science anxiety and self-efficacy. School Psychology Quarterly, 28(4), 360.

Harari, R. R., Vukovic, R. K., & Bailey, S. P. (2013). Mathematics anxiety in young children: An exploratory study. The Journal of Experimental Education, 81(4), 538–555.

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. Journal for Research in Mathematics Education, 33–46.

Hofstede, G. H., Hofstede, G. J., & Minkov, M. (2010). Cultures and organizations: Software of the mind (3rd ed.). Mcgraw-hill New York.

Jackson, C. D., & Leffingwell, R. J. (1999). The role of instructors in creating math anxiety in students from kindergarten through college. The Mathematics Teacher, 92(7), 583–586.

Jameson, M. M. (2014). Contextual factors related to math anxiety in second-grade children. The Journal of Experimental Education, 82(4), 518–536.

Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. Learning and Individual Differences, 19(3), 355–365.

Lu, Y., Li, Q., Patrick, H., & Mantzicopoulos, P. (2019). "Math Gives Me a Tummy Ache!" Mathematics Anxiety in Kindergarten. The Journal of Experimental Education, 1–17.

Maloney, E. A., Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2015). Intergenerational effects of parents' math anxiety on children's math achievement and anxiety. Psychological Science, 26(9), 1480–1488.

Miller, H., & Bichsel, J. (2004). Anxiety, working memory, gender, and math performance. Personality and Individual Differences, 37(3), 591–606.

Novak, E., & Tassell, J. L. (2017). Studying preservice teacher math anxiety and mathematics performance in geometry, word, and non-word problem solving. Learning and Individual Differences, 54, 20–29.

Onslow, B. (1993). Improving the attitude of students and parents through family involvement in mathematics. Mathematics Education Research Journal, 4(3), 24–31.

Parsons, J. E., Adler, T. F., & Kaczala, C. M. (1982). Socialization of achievement attitudes and beliefs: Parental influences. Child Development, 310–321.

Radišić, J., Videnović, M., & Baucal, A. (2015). Math anxiety—Contributing school and individual level factors. European Journal of Psychology of Education, 30(1), 1–20.

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. Journal of Cognition and Development, 14(2), 187–202.

Ramirez, G., Hooper, S. Y., Kersting, N. B., Ferguson, R., & Yeager, D. (2018). Teacher math anxiety relates to adolescent students' math achievement. AERA Open, 4(1), 2332858418756052.

Satake, E., & Amato, P. P. (1995). Mathematics anxiety and achievement among Japanese elementary school students. Educational and Psychological Measurement, 55(6), 1000–1007.

Schaeffer, M. W., Rozek, C. S., Maloney, E. A., Berkowitz, T., Levine, S. C., & Beilock, S. L. (2020). Elementary school teachers' math anxiety and students' math learning: A large‐scale replication. Developmental Science, e13080.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. Review of Educational Research, 75(3), 417–453.

Soni, A., & Kumari, S. (2017). The role of parental math anxiety and math attitude in their children's math achievement. International Journal of Science and Mathematics Education, 15(2), 331–347.

Stipek, D. J., & Ryan, R. H. (1997). Economically disadvantaged preschoolers: Ready to learn but further to go. Developmental Psychology, 33(4), 711.

Supekar, K., Iuculano, T., Chen, L., & Menon, V. (2015). Remediation of childhood math anxiety and associated neural circuits through cognitive tutoring. Journal of Neuroscience, 35(36), 12574–12583.

Woodard, T. (2004). The Effects of Math Anxiety on Post-Secondary Developmental Students as Related to Achievement, Gender, and Age. Inquiry, 9(1), n1.

Yüksel-Şahin, F. (2008). Mathematics anxiety among 4th and 5th grade Turkish elementary school students. International Electronic Journal of Mathematics Education, 3(3), 179–192.

Appendix 8 - Standardized Coefficient and Effect Sizes

Table A8-1- Standardized Coefficient and Effect Sizes PISA

Country	Individual Effect – Standardized Coefficient	Individual Effect – Effect Size	Contextual Effect – Standardized Coefficient	Contextual Effect – Effect Size
Albania	-0.005(0.022)	-0.01(0.045)	0.041(0.067)	0.084(0.136)
Argentina	-0.114(0.016)***	-0.302(0.043)***	-0.145(0.085)	-0.385(0.215)
Australia	-0.213(0.011)***	-0.503(0.024)***	-0.119(0.042)**	-0.282(0.099)**
Austria	-0.176(0.013)***	-0.484(0.036)***	-0.198(0.099)*	-0.549(0.264)*
Belgium	-0.115(0.01)***	-0.338(0.029)***	-0.159(0.075)*	-0.47(0.212)*
Brazil	-0.121(0.011)***	-0.342(0.031)***	-0.128(0.061)*	-0.356(0.163)*
Bulgaria	-0.128(0.015)***	-0.335(0.039)***	-0.246(0.067)***	-0.635(0.159)***
Canada	-0.27(0.016)***	-0.608(0.037)***	-0.085(0.05)	-0.193(0.113)
Chile	-0.132(0.013)***	-0.369(0.036)***	0.021(0.09)	0.06(0.26)
Colombia	-0.133(0.016)***	-0.347(0.04)***	-0.04(0.085)	-0.103(0.221)
Costa Rica	-0.142(0.017)***	-0.393(0.045)***	-0.126(0.1)	-0.351(0.274)
Croatia	-0.187(0.016)***	-0.471(0.038)***	-0.224(0.07)**	-0.56(0.171)**
Czechia	-0.206(0.013)***	-0.547(0.044)***	-0.236(0.11)*	-0.628(0.281)*
Denmark	-0.3(0.016)***	-0.648(0.035)***	-0.188(0.079)*	-0.414(0.173)*
Estonia	-0.267(0.016)***	-0.585(0.033)***	-0.082(0.051)	-0.179(0.111)
Finland	-0.227(0.017)***	-0.474(0.04)***	-0.242(0.162)	-0.543(0.349)
France	-0.13(0.016)***	-0.378(0.045)***	-0.055(0.052)	-0.155(0.147)
Germany	-0.161(0.013)***	-0.458(0.041)***	-0.252(0.074)**	-0.715(0.193)***
Greece	-0.203(0.016)***	-0.49(0.034)***	-0.214(0.062)**	-0.521(0.142)***
Hong Kong	-0.17(0.013)***	-0.405(0.036)***	-0.351(0.098)***	-0.852(0.222)***
Hungary	-0.133(0.015)***	-0.396(0.047)***	-0.28(0.073)***	-0.805(0.184)***
Iceland	-0.27(0.016)***	-0.569(0.031)***	-0.09(0.041)*	-0.189(0.087)*
Indonesia	-0.058(0.013)***	-0.16(0.034)***	-0.054(0.058)	-0.147(0.158)
Ireland	-0.213(0.017)***	-0.474(0.037)***	-0.1(0.042)*	-0.222(0.092)*
Israel	-0.109(0.015)***	-0.282(0.039)***	-0.204(0.092)*	-0.539(0.228)*
Italy	-0.148(0.008)***	-0.422(0.023)***	-0.054(0.066)	-0.16(0.192)
Japan	-0.121(0.013)***	-0.356(0.036)***	0.045(0.101)	0.133(0.301)
Jordan	-0.105(0.02)***	-0.259(0.048)***	-0.159(0.058)**	-0.388(0.141)**
Kazakhstan	-0.131(0.017)***	-0.338(0.042)***	0.029(0.076)	0.076(0.195)
Korea, Republic of	-0.091(0.031)**	-0.226(0.081)**	-0.113(0.091)	-0.274(0.221)

Latvia	-0.218(0.018)***	-0.496(0.04)***	-0.066(0.062)	-0.149(0.138)
Lithuania	-0.224(0.016)***	-0.523(0.039)***	-0.143(0.056)*	-0.332(0.127)**
Luxembourg	-0.177(0.011)***	-0.413(0.022)***	-0.218(0.083)**	-0.508(0.184)**
Macao	-0.192(0.014)***	-0.451(0.037)***	-0.071(0.097)	-0.175(0.237)
Malaysia	-0.109(0.02)***	-0.269(0.047)***	-0.081(0.059)	-0.201(0.147)
Mexico	-0.164(0.01)***	-0.409(0.023)***	-0.03(0.063)	-0.074(0.155)
Montenegro	-0.163(0.016)***	-0.367(0.039)***	-0.347(0.099)***	-0.795(0.218)***
Netherlands	-0.097(0.015)***	-0.318(0.055)***	-0.235(0.089)**	-0.763(0.258)**
New Zealand	-0.195(0.02)***	-0.444(0.048)***	-0.195(0.063)**	-0.446(0.141)**
Norway	-0.3(0.018)***	-0.635(0.047)***	-0.27(0.102)**	-0.588(0.225)**
Perm(Russian Federation)	-0.189(0.027)***	-0.448(0.063)***	-0.142(0.082)	-0.328(0.189)
Peru	-0.115(0.013)***	-0.318(0.036)***	-0.108(0.076)	-0.295(0.205)
Poland	-0.327(0.015)***	-0.71(0.041)***	-0.286(0.099)**	-0.642(0.217)**
Portugal	-0.159(0.014)***	-0.386(0.033)***	-0.115(0.058)*	-0.28(0.138)*
Qatar	-0.132(0.01)***	-0.342(0.026)***	-0.27(0.046)***	-0.695(0.113)***
Romania	-0.118(0.016)***	-0.286(0.049)***	-0.401(0.139)**	-1.016(0.295)**
Russian Federation	-0.211(0.015)***	-0.492(0.034)***	-0.111(0.049)*	-0.258(0.112)*
Serbia	-0.147(0.019)***	-0.38(0.059)***	-0.242(0.185)	-0.629(0.466)
Shanghai-China	-0.183(0.015)***	-0.453(0.036)***	-0.149(0.062)*	-0.365(0.151)*
Singapore	-0.166(0.015)***	-0.381(0.033)***	-0.325(0.039)***	-0.736(0.086)***
Slovakia	-0.208(0.016)***	-0.545(0.041)***	-0.169(0.059)**	-0.417(0.142)**
Slovenia	-0.139(0.011)***	-0.429(0.032)***	0.023(0.081)	0.068(0.24)
Spain	-0.169(0.011)***	-0.376(0.025)***	-0.078(0.026)**	-0.173(0.058)**
Sweden	-0.239(0.016)***	-0.521(0.034)***	0.061(0.124)	0.136(0.278)
Switzerland	-0.187(0.013)***	-0.449(0.03)***	-0.068(0.061)	-0.162(0.144)
Taiwan	-0.124(0.019)***	-0.312(0.05)***	0.055(0.146)	0.141(0.377)
Thailand	-0.053(0.018)**	-0.144(0.05)**	-0.18(0.127)	-0.508(0.338)
Tunisia	-0.033(0.017)*	-0.086(0.044)*	-0.135(0.168)	-0.348(0.434)
Turkey	-0.078(0.017)***	-0.24(0.054)***	-0.31(0.095)**	-0.935(0.242)***
United Arab Emirates	-0.163(0.01)***	-0.404(0.031)***	-0.331(0.063)***	-0.823(0.138)***
United Kingdom	-0.204(0.016)***	-0.472(0.038)***	-0.29(0.047)***	-0.666(0.105)***
United States	-0.259(0.018)***	-0.585(0.044)***	-0.229(0.097)*	-0.528(0.223)*
Uruguay	-0.161(0.016)***	-0.42(0.044)***	-0.202(0.11)	-0.533(0.279)
Viet Nam	-0.115(0.017)***	-0.339(0.052)***	-0.251(0.114)*	-0.696(0.277)*

Note. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Table A8-2- Standardized Coefficient and Effect Sizes TIMSS 4

Country	Individual Effect – Standardized Coefficient	Individual Effect – Effect Size	Contextual Effect – Standardized Coefficient	Contextual Effect – Effect Size
Argentina, Buenos Aires	-0.179(0.022)***	-0.436(0.054)***	-0.128(0.052)*	-0.31(0.123)*
Armenia	-0.246(0.014)***	-0.546(0.03)***	-0.161(0.03)***	-0.356(0.064)***
Australia	-0.236(0.017)***	-0.542(0.039)***	-0.202(0.039)***	-0.465(0.087)***
Bahrain	-0.21(0.013)***	-0.474(0.028)***	-0.155(0.022)***	-0.348(0.05)***
Belgium (Flemish)	-0.178(0.017)***	-0.402(0.038)***	-0.047(0.036)	-0.107(0.082)
Bulgaria	-0.197(0.016)***	-0.483(0.038)***	-0.26(0.047)***	-0.628(0.11)***
Canada	-0.252(0.009)***	-0.584(0.021)***	-0.154(0.03)***	-0.359(0.073)***
Canada (Ontario)	-0.266(0.014)***	-0.593(0.03)***	-0.117(0.037)**	-0.261(0.083)**
Canada (Quebec)	-0.264(0.019)***	-0.582(0.041)***	-0.056(0.043)	-0.124(0.096)
Chile	-0.214(0.012)***	-0.485(0.027)***	-0.319(0.033)***	-0.716(0.071)***
Croatia	-0.266(0.024)***	-0.567(0.049)***	0.022(0.058)	0.048(0.125)
Cyprus	-0.284(0.013)***	-0.6(0.027)***	-0.027(0.022)	-0.056(0.046)
Czechia	-0.301(0.014)***	-0.659(0.028)***	0.022(0.046)	0.049(0.102)
Denmark	-0.231(0.019)***	-0.51(0.04)***	-0.122(0.029)***	-0.267(0.064)***
England	-0.2(0.015)***	-0.473(0.036)***	-0.112(0.041)**	-0.268(0.097)**
Finland	-0.216(0.016)***	-0.458(0.035)***	-0.163(0.039)***	-0.352(0.086)***
France	-0.218(0.015)***	-0.494(0.034)***	-0.083(0.045)	-0.189(0.101)
Georgia	-0.231(0.018)***	-0.532(0.041)***	-0.252(0.057)***	-0.583(0.128)***
Germany	-0.252(0.015)***	-0.553(0.033)***	-0.062(0.044)	-0.138(0.097)
Hong Kong	-0.237(0.018)***	-0.559(0.037)***	-0.202(0.035)***	-0.482(0.085)***
Hungary	-0.208(0.015)***	-0.504(0.036)***	-0.178(0.048)***	-0.431(0.117)***
Indonesia	-0.184(0.017)***	-0.471(0.047)***	-0.256(0.037)***	-0.644(0.086)***
Iran, Islamic Republic of	-0.285(0.021)***	-0.695(0.049)***	-0.135(0.056)*	-0.325(0.132)*
Ireland	-0.309(0.018)***	-0.658(0.037)***	0(0.049)	0.001(0.104)
Italy	-0.18(0.018)***	-0.398(0.039)***	0.036(0.036)	0.08(0.081)
Japan	-0.227(0.016)***	-0.469(0.032)***	-0.015(0.021)	-0.03(0.044)
Kazakhstan	-0.122(0.015)***	-0.354(0.045)***	-0.215(0.054)***	-0.608(0.142)***
Korea, Republic of	-0.171(0.014)***	-0.367(0.03)***	-0.057(0.045)	-0.122(0.097)

Kuwait	-0.182(0.023)***	-0.434(0.054)***	-0.065(0.026)*	-0.154(0.062)*
Lithuania	-0.241(0.018)***	-0.554(0.039)***	-0.148(0.05)**	-0.341(0.112)**
Morocco	-0.186(0.011)***	-0.488(0.029)***	-0.132(0.039)**	-0.35(0.102)**
Netherlands	-0.216(0.018)***	-0.456(0.038)***	-0.089(0.037)*	-0.188(0.078)*
New Zealand	-0.204(0.013)***	-0.483(0.03)***	-0.142(0.031)***	-0.336(0.072)***
Northern Ireland	-0.296(0.021)***	-0.644(0.043)***	-0.146(0.043)**	-0.316(0.093)**
Norway	-0.253(0.016)***	-0.535(0.032)***	-0.031(0.028)	-0.065(0.059)
Norway (4)	-0.257(0.016)***	-0.547(0.036)***	-0.094(0.028)**	-0.198(0.06)**
Oman	-0.197(0.011)***	-0.443(0.024)***	-0.13(0.025)***	-0.291(0.056)***
Poland	-0.271(0.015)***	-0.587(0.031)***	0.008(0.022)	0.017(0.048)
Portugal	-0.281(0.015)***	-0.632(0.032)***	-0.128(0.036)***	-0.286(0.081)***
Qatar	-0.184(0.013)***	-0.461(0.033)***	-0.166(0.037)***	-0.415(0.092)***
Russian Federation	-0.24(0.017)***	-0.597(0.041)***	-0.193(0.041)***	-0.475(0.1)***
Saudi Arabia	-0.161(0.012)***	-0.392(0.031)***	-0.218(0.04)***	-0.534(0.095)***
Serbia	-0.283(0.02)***	-0.599(0.042)***	0.227(0.049)***	0.495(0.107)***
Singapore	-0.133(0.008)***	-0.351(0.022)***	-0.435(0.028)***	-1.127(0.06)***
Slovakia	-0.255(0.014)***	-0.622(0.032)***	-0.037(0.048)	-0.092(0.119)
Slovenia	-0.262(0.018)***	-0.544(0.038)***	0.024(0.033)	0.05(0.069)
Spain	-0.282(0.013)***	-0.629(0.027)***	-0.04(0.03)	-0.09(0.068)
Sweden	-0.211(0.018)***	-0.469(0.041)***	-0.17(0.038)***	-0.379(0.083)***
Taiwan	-0.261(0.014)***	-0.544(0.028)***	-0.07(0.037)	-0.146(0.076)
Turkey	-0.296(0.012)***	-0.698(0.032)***	-0.324(0.037)***	-0.738(0.083)***
United Arab Emirates	-0.15(0.006)***	-0.406(0.018)***	-0.302(0.022)***	-0.816(0.053)***
United Arab Emirates (Abu Dhabi)	-0.153(0.01)***	-0.397(0.03)***	-0.374(0.039)***	-0.974(0.086)***
United Arab Emirates (Dubai)	-0.152(0.008)***	-0.412(0.022)***	-0.239(0.033)***	-0.648(0.085)***
United States	-0.226(0.009)***	-0.544(0.023)***	-0.235(0.029)***	-0.554(0.067)***

Note. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Table A8-3- Standardized Coefficient and Effect Sizes TIMSS 8

Country	Individual Effect – Standardized Coefficient	Individual Effect – Effect Size	Contextual Effect – Standardized Coefficient	Contextual Effect – Effect Size
Argentina, Buenos Aires	-0.166(0.017)***	-0.436(0.044)***	-0.048(0.068)	-0.127(0.179)
Armenia	-0.168(0.017)***	-0.377(0.037)***	0.048(0.067)	0.11(0.153)
Australia	-0.183(0.012)***	-0.521(0.034)***	-0.185(0.039)***	-0.537(0.11)***
Bahrain	-0.159(0.02)***	-0.369(0.045)***	-0.088(0.059)	-0.215(0.147)
Botswana	-0.114(0.015)***	-0.247(0.033)***	-0.119(0.032)***	-0.258(0.07)***
Canada	-0.29(0.013)***	-0.701(0.029)***	-0.047(0.037)	-0.116(0.091)
Canada (Ontario)	-0.295(0.016)***	-0.675(0.037)***	-0.164(0.039)***	-0.381(0.091)***
Canada (Quebec)	-0.274(0.018)***	-0.711(0.044)***	-0.081(0.074)	-0.21(0.191)
Chile	-0.169(0.015)***	-0.424(0.039)***	-0.136(0.051)**	-0.34(0.126)**
Egypt	-0.147(0.012)***	-0.345(0.029)***	-0.145(0.042)**	-0.342(0.098)***
England	-0.094(0.014)***	-0.329(0.053)***	-0.35(0.057)***	-1.262(0.133)***
Georgia	-0.219(0.019)***	-0.491(0.044)***	-0.177(0.046)***	-0.394(0.102)***
Hong Kong	-0.19(0.013)***	-0.562(0.038)***	-0.089(0.07)	-0.276(0.212)
Hungary	-0.181(0.014)***	-0.495(0.039)***	-0.168(0.049)**	-0.457(0.134)**
Iran, Islamic Republic of	-0.187(0.012)***	-0.49(0.031)***	-0.158(0.045)***	-0.403(0.114)***
Ireland	-0.178(0.013)***	-0.505(0.036)***	-0.06(0.038)	-0.172(0.107)
Israel	-0.088(0.01)***	-0.298(0.032)***	-0.121(0.033)***	-0.397(0.106)***
Italy	-0.273(0.019)***	-0.624(0.038)***	-0.023(0.034)	-0.052(0.078)
Japan	-0.17(0.012)***	-0.384(0.028)***	0.018(0.047)	0.04(0.107)
Jordan	-0.164(0.014)***	-0.375(0.032)***	-0.046(0.046)	-0.105(0.105)
Kazakhstan	-0.109(0.014)***	-0.333(0.041)***	-0.049(0.054)	-0.149(0.164)
Korea, Republic of	-0.062(0.016)***	-0.144(0.037)***	0.07(0.05)	0.174(0.123)
Kuwait	-0.102(0.02)***	-0.264(0.051)***	-0.012(0.071)	-0.033(0.186)
Lebanon	-0.115(0.02)***	-0.288(0.05)***	-0.075(0.056)	-0.184(0.135)
Lithuania	-0.25(0.017)***	-0.57(0.041)***	-0.141(0.052)**	-0.321(0.119)**
Malaysia	-0.142(0.01)***	-0.401(0.031)***	-0.404(0.044)***	-1.139(0.089)***
Malta	-0.153(0.012)***	-0.515(0.039)***	0.011(0.065)	0.041(0.236)
Morocco	-0.163(0.012)***	-0.367(0.026)***	-0.068(0.034)*	-0.156(0.078)*
New Zealand	-0.155(0.013)***	-0.458(0.04)***	-0.203(0.04)***	-0.599(0.11)***
Norway	-0.322(0.016)***	-0.678(0.033)***	-0.067(0.031)*	-0.14(0.065)*

Norway (8)	-0.322(0.017)***	-0.683(0.037)***	-0.102(0.037)**	-0.214(0.077)**
Oman	-0.173(0.014)***	-0.388(0.031)***	-0.05(0.034)	-0.112(0.075)
Qatar	-0.163(0.014)***	-0.432(0.039)***	-0.225(0.052)***	-0.606(0.136)***
Russian Federation	-0.234(0.013)***	-0.582(0.032)***	-0.089(0.055)	-0.219(0.134)
Saudi Arabia	-0.11(0.02)***	-0.258(0.045)***	-0.089(0.049)	-0.21(0.119)
Singapore	-0.127(0.008)***	-0.4(0.031)***	-0.434(0.039)***	-1.35(0.076)***
Slovenia	-0.234(0.019)***	-0.522(0.039)***	-0.046(0.029)	-0.103(0.066)
South Africa	-0.102(0.01)***	-0.309(0.031)***	-0.105(0.083)	-0.333(0.261)
Sweden	-0.273(0.016)***	-0.625(0.036)***	-0.143(0.061)*	-0.33(0.139)*
Taiwan	-0.199(0.016)***	-0.448(0.036)***	-0.022(0.055)	-0.049(0.124)
Thailand	-0.113(0.012)***	-0.317(0.036)***	-0.236(0.065)***	-0.651(0.174)***
Turkey	-0.262(0.013)***	-0.622(0.031)***	-0.25(0.049)***	-0.583(0.113)***
United Arab Emirates	-0.179(0.009)***	-0.518(0.025)***	-0.052(0.036)	-0.151(0.104)
United Arab Emirates (Abu Dhabi)	-0.167(0.012)***	-0.467(0.032)***	-0.132(0.071)	-0.375(0.199)
United Arab Emirates (Dubai)	-0.212(0.02)***	-0.592(0.055)***	-0.057(0.048)	-0.161(0.137)
United States	-0.154(0.007)***	-0.445(0.023)***	-0.262(0.032)***	-0.752(0.083)***

Note. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Appendix 9 - Additional Analyses

Is there a Contextual Effect of Math Anxiety at the Country Level?

Description

Just as students are nested in education environment, education environments are nested into different countries. This additional level of nesting raises the possibility that between-country differences in math anxiety may predict student achievement. To investigate whether country-average math anxiety may affect student achievement, we estimated a simple three-level model for each database. This model will yield estimates of the individual effect, the education environment contextual effect and the country contextual effect. Specifically, we have the following model:

$$\text{L1: } Y_{i,j,k} = \pi_{0,j,k} + \pi_{1,j,k}\left(X_{i,j,k} - \bar{X}_{\cdot,\cdot,\cdot}\right) + e_{i,j,k}$$

$$\text{L2: } \pi_{0,j,k} = \beta_{0,0,k} + \beta_{0,1,k}\left(X_{\cdot,j,k} - \bar{X}_{\cdot,\cdot,\cdot}\right) + r_{0,j,k}$$

$$\text{L3: } \beta_{0,0,k} = \gamma_{0,0,0} + \gamma_{0,0,1}\left(X_{\cdot,\cdot,k} - \bar{X}_{\cdot,\cdot,\cdot}\right) + \mu_{0,0,k}$$

Math anxiety is grand-mean centered, and a manifest aggregation approach is utilized. As such, the higher-level regression coefficients are a direct estimation of the contextual effect that controls for lower-level variations (Brincks et al., 2017; Enders & Tofighi, 2007).

Results

Having found statistically significant between-country differences, we next ascertain whether country-average math anxiety may affect math achievement. We estimated a three-level model for each database. In the TIMSS Grade 4 sample, L1 math anxiety was statistically significant and negatively related to math achievement ($\pi_{1,j,k} = -.172, SE = .005, \Delta = .508$), the contextual effect of L2 math anxiety on math achievement was also statistically significant and negative ($\beta_{0,0,k} = -.152, SE = .019, \Delta = .441$), however, the contextual effect of L3 math anxiety on math achievement was not statistically significant ($\pi_{1,j,k} = -.022, SE = .084, \Delta = .064$).

In the TIMSS Grade 8 sample, L1 math anxiety was statistically significant and negatively related to math achievement ($\pi_{1,j,k} = -.139, SE = .007, \Delta = .450$), the contextual effect of L2 math anxiety on math achievement was also statistically significant and negative ($\beta_{0,0,k} = -.109, SE = .015, \Delta = .349$), however, the contextual effect of L3 math anxiety on math achievement was not statistically significant ($\pi_{1,j,k} = -.009, SE = .030, \Delta = .030$).

In the PISA sample, L1 math anxiety was statistically significant and negatively related to math achievement ($\pi_{1,j,k} = -.146, SE = .006, \Delta = .404$), the contextual effect of L2 math anxiety on math achievement was also statistically significant and negative ($\beta_{0,0,k} = -.131, SE = .018, \Delta = .355$). Finally, the contextual effect of L3 math anxiety on math achievement was statistically significant ($\pi_{1,j,k} = -.229, SE = .046, \Delta = .624$).

In sum, results for L1 and L2 are consistent with the previous analyses – specifically, that there is a negative association between individual math anxiety and math achievement and a negative association between education environment-average math anxiety and math achievement. Consistent with previous analysis, the effects had a small to medium effect size (.349 – .508). Interestingly, the L3 contextual effect that associates between country-average math anxiety and math achievement was statistically significant for the PISA database but were not statistically significant for the TIMSS databases. In other words, results from the PISA database supported the notion that a country's average level of math anxiety is related to children's math achievement but results from the TIMSS databases did not support this hypothesis.

Are there Between-Country Variations in the Magnitude of the Individual and Contextual Effects?

Description

Another possible area of between-country differences in the relations between math anxiety and math achievement is between-country differences in the magnitude of the L1 individual effect and L2 contextual effect. To examine this, we modified the simple three-level model above to allow for the L1 and L2 effects to vary at the higher levels:

$$\text{L1: } Y_{i,j,k} = \pi_{0,j,k} + \pi_{1,j,k}(X_{i,j,k} - \bar{X}_{\cdot,\cdot,\cdot}) + e_{i,j,k}$$

$$\text{L2: } \pi_{0,j,k} = \beta_{0,0,k} + \beta_{0,1,k}(X_{\cdot,j,k} - \bar{X}_{\cdot,\cdot,\cdot}) + r_{0,j,k}$$

$$\pi_{1,j,k} = \beta_{1,0,k} + \mathrm{r}_{1,j,k}$$

$$\text{L3: } \beta_{0,0,k} = \gamma_{0,0,0} + \gamma_{0,0,1}(X_{\cdot,\cdot,k} - \bar{X}_{\cdot,\cdot,\cdot}) + \mu_{0,0,k}$$

$$\beta_{0,1,k} = \gamma_{0,1,0} + u_{0,1,k}$$

$$\beta_{1,0,k} = \gamma_{1,0,0} + \mathrm{u}_{1,0,k}$$

A significant L1 random slope at L2 ($\mathrm{r}_{1,j,k}$) or L3 ($\mathrm{u}_{1,0,k}$) would suggest that the magnitude of relation between individual math anxiety and math achievement differ as a function of membership in the education environment or country, or both. Relatedly, a significant L2 random slope at L3 ($u_{0,1,k}$) would suggest the relation between environment-average math anxiety and math achievement differ as a function of membership in country.

Results

We have shown in the previous analyses that there are between-country differences in the magnitude of the L1 individual effects and L2 contextual effects. We next quantify these between-country differences by estimating a three-level model with random-slopes. Results are presented in **Table A9-1**.

Results indicate that magnitude of effect of individual math anxiety on math achievement ($\gamma_{1,0,0}$) to significantly differ at L2 ($\mathrm{r}_{1,j,k}$) and L3 ($\mathrm{u}_{1,0,k}$) for all three databases. This suggest that the effect that individual math anxiety has on math achievement differ depending on the education environment and country membership. There was a significant average L1 individual effect of math anxiety on math achievement for the TIMSS Grade 4 ($\gamma_{1,0,0} = -.179, SE = .005, \Delta = .531$), TIMSS Grade 8 ($\gamma_{1,0,0} = -.143, SE = .007, \Delta = .462$), and PISA ($\gamma_{1,0,0} = -.139, SE = .007, \Delta = .383$). To ascertain the typical range of values that the individual effects may differ between countries and education environments, we calculated a 95% plausible value range (Raudenbush & Bryk, 2002), which theoretically encompasses the magnitude of the individual effect for 95% of education environments and countries (Lorah, 2018). The 95% plausible value range was -.286 to -.072, -.267 to -.019 and -.291 to .013 for TIMSS Grade 4, TIMSS Grade 8 and PISA, respectively.

Similarly, we find that the magnitude of education environment-average math anxiety on math achievement ($\gamma_{0,1,0}$) to significantly differ at L3 ($u_{0,1,k}$) for all three databases. This suggest that the effect that environment-average math anxiety has on math achievement differ depending on country membership. There was a significant average L2 contextual effect for the TIMSS Grade 4 ($\gamma_{0,1,0} = -.117, SE = .014, \Delta = .382$), TIMSS Grade 8 ($\gamma_{0,1,0} = -.102, SE = .015, \Delta = .395$), and PISA ($\gamma_{0,1,0} = -.129, SE = .013, \Delta = .466$). The 95% plausible value range, the degree of variability between countries for the magnitude of the contextual effect was -.315 to .081, -.290 to .085 and -.317 to .059 for TIMSS Grade 4, TIMSS Grade 8 and PISA, respectively.

Finally, we find that the effect that country-average math anxiety has on math achievement ($\gamma_{0,0,1}$) was insignificant for the TIMSS Grade 4 ($\gamma_{0,0,1} = -.018, SE = .077, \Delta = .058$), and Grade 8 ($\gamma_{0,0,1} = -.009, SE = .076, \Delta = .021$), but was statistically significant for the PISA ($\gamma_{0,0,1} = -.215, SE = .042, \Delta = .604$).

In sum, results from the three-level model with random slopes revealed that on average, there is a L1 individual effect of math anxiety on math achievement. Extending the previous results, we find that both the education environment membership and country membership contributes to variability in how individual's math anxiety may affect math achievement. Similarly, we find that on average there is a L2 contextual effect of math anxiety on math achievement. Further, we find that contextual effect depends on country membership.

Table A9-1

Results for the Three-Level Model with Random-Slopes

	TIMSS Grade 4	*TIMSS Grade 8*	*PISA*
	Coefficient (SE)	*Coefficient (SE)*	*Coefficient (SE)*
Fixed Effect			
Average Initial Math Achievement, $\gamma_{0,0,0}$	.047(.084)	.022(.088)	-.101(.054)
Individual Math Anxiety, $\gamma_{1,0,0}$	-.179(.005)***	-.143(.007)***	-.139(.007)***
Education Environment Math Anxiety, $\gamma_{0,1,0}$	-.117(.014)***	-.102(.015)***	-.129(.013)***
Country Math Anxiety, $\gamma_{0,0,1}$	-.018(.077)	-.009(.076)	-.215(.042)***
Random Effect			
Math Achievement L1 Residual, $e_{i,j,k}$	.406(.015)***	.348(.020)***	.430(.017)***
Math Achievement L2 Residual, $r_{0,j,k}$	.189(.025)***	.264(.022)***	.261(.019)***
Math Achievement L3 Residual, $\mu_{0,0,k}$	.322(.070)***	.330(.071)***	.175(.037)***
Individual Math Anxiety at L2, $r_{1,j,k}$	.002(.001)**	.002(.000)***	.003(.001)***
Individual Math Anxiety at L3, $u_{1,0,k}$	.001(.000)***	.002(.000)***	.003(.000)***
Education Environment Math Anxiety at L3, $u_{0,1,k}$	.010(.003)***	.009(.003)**	.009(.002)***
Average Effect Size (Δ)			
Individual Math Anxiety	.531	.462	.383
Education Environment Math Anxiety	.382	.395	.466
Country Math Anxiety	.058	.021	.604

Note. ***$p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Curriculum Vitae

Name: Nathan T.T. Lau

Post-secondary Education and Degrees:

2018-2021	PhD Cognitive, Developmental, and Brain Sciences
	University of Western Ontario
	Advisor: Daniel Ansari
2017-2018	Master of Philosophy in Psychology
	University of Hong Kong
2006-2010	Bachelor of Commerce
	McMaster University

Honours and Awards:

2018-2021	Ontario Graduate Scholarship

Related Work Experience:

2020-2021	Teaching Assistant for Statistics for Psychology
	University of Western Ontario
2019	Teaching Assistant for Psychology of Perception
	University of Western Ontario
2018	Teaching Assistant for Introduction to Psychology
	University of Western Ontario
2017	Teaching Assistant for Cognitive Psychology
	University of Hong Kong

Publications:

Lau, N. T., Merkley, R., Tremblay, P., Zhang, S., De Jesus, S., & Ansari, D. (2021). Kindergarteners' symbolic number abilities predict nonsymbolic number abilities and math achievement in grade 1. *Developmental Psychology*.

Chan, W. W. L., Au, T. K., Lau, N. T., & Tang, J. (2017). Counting errors as a window onto children's place-value concept. *Contemporary Educational Psychology*.