UMEÅ UNIVERSITY

Math Anxiety in Primary School Students

Measurement, Mediators, and Cross-Cultural Comparisons

Jonatan Finell

Department of Applied Educational Science
Umeå 2025

Academic Dissertations in Beteendevetenskapliga Mätningar No. 16
ISBN: 978-91-8070-676-6 (print)
ISBN: 978-91-8070-677-3 (pdf)
ISSN: 1652-9650
Cover art and design by: Björn Sigurdsson
Printed by: Scandinavian Print Group
Umeå, Sweden 2025

Fact and fiction–
they are not always easy to distinguish.

- Monsignor Quixote

Table of Contents

Abstract

This thesis was conducted within the project *Choking Under Pressure*, a longitudinal study following Swedish Grade 4 students during four semesters. The main purpose of this thesis was to develop and validate a math anxiety rating scale for Swedish-speaking primary school students in a Nordic context, and to explore the underlying mechanisms in the math anxiety–math performance relationship. This thesis consists of four studies, where Studies I and II adopt a measurement perspective. Study III provides a review of the role of working memory in the math anxiety–math performance relationship. Study IV extends Study III by empirically examining working memory and self-concept as mediators.

Study I assessed the validity and reliability of the Swedish MARS-E. Test-retest reliability and internal consistency demonstrated longitudinal stability and strong item agreement. Dimensionality analysis supported a one-factor structure, with evidence of both gender and longitudinal invariance. Significant gender differences in math anxiety were found, increasing across time points.

Study II extended Study I by combining data from two Finnish projects–one Finnish and one Finland-Swedish–with the Swedish sample. All measurements were conducted in Spring 2023 on Grade 4 students using the same instruments. A one-factor model was found for both Swedish-speaking groups, whereas the Finnish version of the MARS-E supported a two-factor model.

Study III was a literature review employing meta-analysis to examine the relationship between working memory and math anxiety, and the mediating role of working memory in the math anxiety–performance link. Results from 57 studies showed a significant negative correlation, and data from 8 studies indicated a partial mediation effect.

Study IV empirically examined how working memory and math self-concept influence the math anxiety–math performance link. Using structural equation models across two longitudinal waves, as well as multiple mediation, the results showed that both working memory and self-concept in math are important mediators in this relationship.

The main contributions of this thesis are the development and validation of the Swedish MARS-E, as well as a deeper understanding of the mediating roles of working memory and math self-concept in the math anxiety–math performance relationship. These findings have implications for both research and educational practice, refining theoretical models of math anxiety and its cognitive and motivational correlates, and providing a foundation for interventions targeting math anxiety.

Populärvetenskaplig sammanfattning

Matematikkunskap är avgörande för att funktionera på en hög nivå i dagens samhälle, det gäller såväl skola som arbetsliv och vardag. En faktor som hämmar inlärning och prestationer i matematik är matematikängslan. Det här är särskilt oroande då många elever upplever matematikängslan redan under tidiga skolår, och dessutom ökar prevalensen med ålder. PISA 2022 mätningen visar att 25 % av svenska 15-åringar upplever medel till hög grad av matematikängslan, motsvarande siffra för finländska elever är 20 %. Matematikängslan är förknippad med elevens oro och nervositet inför matematiska aktiviteter, och ger sig uttryck i både negativa tankar och fysiologiska reaktioner. En central aspekt av matematikängslan är rädslan inför att ta sig an en matematikuppgift.

Den här avhandlingen har utarbetats inom ramen för ett longitudinellt projekt ”Att inte klara av trycket”. Kvantitativa data från 18 skolor i Sverige samlades in under fyra tidpunkter. Avhandlingen består av fyra delstudier som tillsammans:

1. Redogör för valideringen av ett nytt instrument i form av en enkät för att mäta matematikängslan hos svensktalande grundskoleelever, och
2. Redogör för vilka underliggande mekanismer som kan förklara hur matematikängslan påverkar matematikprestationer.

Studie I beskriver hur valideringen av instrumentet MARS-E genomfördes, en enkät som vi i projektgruppen utvecklade i syfte att mäta matematikängslan. Ca 400 elever i årskurs 4, utspridda över sex kommuner i Sverige besvarade enkäten samt testades med ett batteri av matematiktester. Resultaten visade att instrumentet mäter ett endimensionellt konstrukt av generell matematikängslan. Genom en serie av konfirmatoriska faktoranalyser kunde vi även säkra att vi mäter samma konstrukt över tid och mellan pojkar och flickor. Studien bekräftar att matematikängslan är nära besläktad med, men separat från testängslan. Matematikängslan visade sig vara förknippad med lägre prestationer i matematik, samt betydligt sämre självuppfattning i matematik.

Studie II kan ses som en fortsättning på Studie I, där vi slog samman data från två systerprojekt i Finland–ett finskt och ett finlandssvenskt–som utförde samma slags mätningar på elever från samma årskurser. Resultaten från studien bekräftar att MARS-E instrumentet fungerar ekvivalent över de två svensktalande grupperna (sverigesvenska – finlandssvenska). Bland de svensktalande eleverna visar resultaten att matematikängslan främst utgörs av en enda övergripande dimension. Hos de finsktalande eleverna däremot framträder två olika delar av ängslan: en kognitiv del, som handlar om tankar och oro kring prestation, och en affektiv del, som rör känslomässiga reaktioner som stress eller obehag. Studien bekräftar även tidigare resultat gällande könsskillnader, där pojkar rapporterar lägre nivåer av matematikängslan jämfört med flickor. Vi fann även att pojkars matematikprestationer påverkas mera av matematikängslan än vad flickors gör.

Studie III genomfördes som en systematisk litteraturstudie med metaanalys som metod för analysen. Syftet med Studie III var att sammanställa alla publicerade effektstorlekar som mäter associationen mellan arbetsminne och matematikängslan, samt sammanställa den indirekta effekten av matematikprestationers påverkan på matematik när man beaktar arbetsminne som en potentiell medlare. Utifrån 57 publicerade studier fann vi ett svagt, men signifikant, negativt samband mellan matematikängslan och arbetsminne. Arbetsminne visade sig även åtminstone delvis förklara en del av variationen i matematikprestationer som annars härstammar från matematikängslan.

Syftet med Studie IV var att empiriskt testa hur arbetsminne och självuppfattning i matematik, mätt genom latenta variabler, kan mediera framtida matematikprestationer, samt hur könsspecifika effekter skiljer sig. Resultaten visade att både arbetsminne och självuppfattning åtminstone partiellt kan mediera den negativa effekten på matematikprestationer som har sitt ursprung i matematikängslan. När både arbetsminne och självuppfattning tas i beaktande förklarar de tillsammans den variation som matematikängslan annars hade förklarat.

Den här avhandlingen fördjupar vår förståelse av konceptet matematikängslan och dess roll i elevernas matematiska utveckling. Genom olika forskningsmetoder visar arbetet att matematikängslan är ett mätbart konstrukt med konsekvenser för arbetsminne, självuppfattning och matematikprestationer. Utöver att bidra till teoretiska modeller lyfter den här forskningen fram vikten av tidig identifiering och insats. Att hantera matematikängslan handlar inte enbart om att minska oro, negativa känslor eller rädsla, utan också om att främja lärande, stärka akademisk utveckling på lång sikt och potentiellt påverka framtida utbildnings- och yrkesval.

Acknowledgements

Pursuing a PhD is a somewhat paradoxical journey. Much of the work is done in solitude, with hours spent reading and writing alone. Yet, this kind of work cannot be completed alone. It is shaped and supported by many remarkable people along the way. This is why a thank you is in order.

Above all else, I want to extend my gratitude to my supervisors Bert, Hanna, and Johan. Bert, my main supervisor, has always been the go-to person for problems, no matter how big or small. Hanna made sure I stayed on track with my studies, and her feedback—especially on the kappa—was invaluable. Johan was the one who first inspired me to pursue a PhD and has been providing continuous advice for a decade now. I feel truly fortunate to have had the opportunity to work closely with this team. The guidance and trust they placed in me have led to some productive moments over the years.

Thank you to my designated readers, Erika and Anna, who read my work and provided extensive feedback. This was truly valuable. And speaking of valuable lessons, the courses provided by the QRM network have shaped me as a researcher and provided the necessary competencies for conducting my studies, I hope future cohorts receive the same opportunity to take these courses.

Thank you to Ellen for your collaboration over these years—I believe we've shown that meta-analysis studies can be fun. Thank you, Pinja, for wrestling with Study II right until the end. Thanks also to the project's research assistant, William, for helping with the extensive data collection, and of course for bringing *fika* on our work trips. Thank you, Björn, for creating the cover for my thesis.

I am also grateful for both my previous and current PhD and post-doc colleagues, as well as the many researchers, teachers, administrative staff, and others at the Department of Applied Educational Science. These people have coloured my day-to-day work and created an enjoyable work environment.

Finally, a salute to my friends and family from Jeppis and the surrounding areas. Thanks to my parents, Gian and Anne, who also offered to proofread from time to time, and to my siblings—Benjamin and Laura together with her family. Lastly, thank you, Erica, for the never-ending support—and thank you Enéa for keeping me grounded no matter the situation.

Umeå, April 2025
Jonatan

Studies

The current thesis includes 4 studies, all of which are listed below. Author contributions for each study are stated in the footnote[1].

Study I	Finell, J., Eklöf, H., Korhonen, J., & Jonsson, B. (2024). Reliability and validity evidence of the Swedish shortened mathematics anxiety rating scale elementary (MARS-E). *Discover Education 3, 240*. https://doi.org/10.1007/s44217-024-00348-8
Study II	Tähti, P., Finell, J., Tapola, A., Sammallahti, E., Widlund, A., Jonsson, B., Mononen, R., & Korhonen, J. (2025). Math anxiety and its relations to arithmetic fluency and number processing: A comparative study between Finnish, Finnish-Swedish, and Swedish fourth-grade students. Submitted for publication. https://doi.org/10.31234/osf.io/r4td6_v1
Study III	Finell, J., Sammallahti, E., Korhonen, J., Eklöf, H., & Jonsson, B. (2022). Working Memory and its mediating role on the relationship of math anxiety and math performance: A meta-analysis. *Frontiers in psychology*, *12*, 798090. https://doi.org/10.3389/fpsyg.2021.798090
Study IV	Finell, J., Jonsson, B., Eklöf, H., Korhonen, J. (2025). Working memory and self-beliefs mediate the anxiety - performance link. Manuscript.

[1] **Study I**: BJ, HE, JK and JF contributed to the study's design. JF collected data, performed the statistical analysis and wrote the first draft of the manuscript. All authors critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.
Study II: Study was conceptualized under joint project meetings. PT wrote the first draft of the manuscript. JF performed the statistical analyses. All authors critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.
Study III: JK and BJ came up with the idea for the study. JF, ES, JK, HE, and BJ jointly contributed to the study's conceptualization and revised the manuscript for important intellectual content. JF collected the data, performed the statistical analysis, and wrote the first draft of the manuscript. All authors contributed to the manuscript and read and approved the submitted version.
Study IV: BJ, HE, JK and JF contributed to the study's design. JF collected data, performed the statistical analysis and wrote the first draft of the manuscript. All authors critically revised the manuscript for important intellectual content.

1 Introduction

This thesis investigates how math anxiety, working memory, and math self-concept relate to math performance among primary school students, in both cross-sectional and longitudinal settings. Central to this work is the measurement and validity of the psychological construct of math anxiety, as valid measurement is a prerequisite for drawing sound inferences.

Math anxiety is a recurring construct in the four studies that are summarised in this thesis. It can act as a significant psychological barrier for both children and adults, often manifesting as negative emotions such as fear or apprehension, and physiological reactions to math-related activities, both in academic situations and everyday life (Ashcraft, 2002; Richardson & Suinn, 1972). However, at the heart of math anxiety lies the fear of failure, which acts as a constant threat to a person's self-esteem (Orbach et al., 2019). Math anxiety tends to be linked with lower math performance, as well as avoidance behaviours (Hembree, 1990), which, in the long run, can shape a person's educational and career choices (Lau et al., 2024). The fact that math anxiety can impact mundane, everyday tasks (such as double-checking a receipt), as well as key choices in life, is cause for concern, especially as studies have consistently reported high prevalence, particularly among adolescent populations (Ramirez et al., 2018). A related construct, math self-concept, refers to how students perceive their own abilities in math (Marsh et al., 1983). These self-beliefs are closely linked to academic achievements, as well as other cognitive outcomes and anxieties (Marsh & Martin, 2011), playing a crucial role in the investigation of the math anxiety–math performance link.

Learning math involves holding numbers in one's mind while performing operations, updating them, and keeping track of several steps at once. This process is cognitively demanding and largely relies on a construct known as working memory. Working memory is a core cognitive function that plays a crucial role in learning mathematics, particularly with large arithmetical problems (Ashcraft & Krause, 2007), which are typical tasks in the higher grades. However, the importance of working memory is not limited to adolescents or complex mathematical tasks. Studies have found clear relationships between working memory and basic mathematics in students as young as seven (Alloway & Passolunghi, 2011). In addition, its predictive power for future math achievement in young children exceeds that of fluid intelligence (Alloway & Alloway, 2010).

Imagine a student trying to solve 16 + 17 in their head. One strategy might be to hold the 6 and 7 in mind, add them to get 13, keep the 3, carry over the 1, and continue the calculation. This kind of task relies on working memory, as the student must both store and actively process information at the same time. These operations are coordinated by the central executive, which manages attention and allocates cognitive resources (Hitch & Baddeley, 1976). This differs from a short-term memory task, where the goal is simply to remember information, like a string of numbers, without needing to manipulate it or manage distractions. Since working memory has a limited capacity, it is important how that capacity is used efficiently. If it is taken up by irrelevant distractions, such as worrying thoughts, there is less available for completing the task effectively. This presents a disadvantage to the distracted individual—or, as this thesis will explore, to the math-anxious individual.

Although there is agreement on the idea that math anxiety is linked to performance, the underlying mechanisms of that connection are still not fully understood. Further research is needed to establish an evidence-based understanding of the relationship. As previously stated, obtaining accurate results require accurate measurement. Since the 1970s, the research community has published many self-report questionnaires designed to measure math anxiety. These target both adults and children, and include various operationalisations of math anxiety, such as anxiety related to learning, evaluation, and affective responses. To my knowledge, there are no formally validated instruments measuring math anxiety targeting Swedish-speaking primary school children. Furthermore, most research on math anxiety has been carried out on older participants. Studying older participants has certain advantages, such as the relative ease of recruiting participants and the fact that math anxiety tends to increase with age (Hembree, 1990), making it more detectable. However, this highlights the importance of studying younger participants. Generalisability is limited across age groups if study results are disproportionate between age groups, and our understanding of the emergence of math anxiety and how it takes form in early development remains incomplete.

1.1 Aims and Research Questions

The main purpose of this thesis was to develop and validate a math anxiety rating scale for Swedish-speaking primary school students in a Nordic context, and to explore the underlying mechanisms in the relationship between math anxiety and math performance. Therefore, the following research questions were formulated:

1. Is the Swedish MARS-E a reliable and valid instrument for measuring math anxiety in Swedish-speaking primary school students?
2. Does the Swedish MARS-E function equivalently across three culturally distinct Nordic samples (Swedish, Finland-Swedish, and Finnish students)?
3. What does existing research reveal about the role of working memory in the math anxiety–math performance relationship?
4. How do working memory and math self-concept mediate the math anxiety–math performance relationship in primary school students?

1.2 Thesis Disposition

This thesis consists of five main chapters. The first chapter introduces the reader to the topic and briefly mentions the four studies that form the basis of the current work. The second chapter goes on to discuss some of the central constructs that appear in the four studies along with theoretical frameworks that have guided the studies and assessment of the Swedish MARS-E's validity evidence. The third chapter provides a description of the main methods used in the four studies. The fourth chapter provides an overview and summary of each study and its main findings. Finally, the fifth chapter reflects on the findings of the studies.

2 Background

This chapter outlines the broader context in which the present work is grounded. The relationship between math anxiety, working memory, math self-concept, and math performance is explored using a variety of methods across the four studies included in this thesis. A key focus of this work is the Swedish Math Anxiety Rating Scale–Elementary (MARS-E), which is formally validated and subsequently used to measure math anxiety in relation to performance. Two theoretical models, the Attentional Control Theory and the Debilitating Anxiety Model, help frame this work and explain the cognitive and emotional processes that underlie the math anxiety–math performance relationship. This chapter provides an overview of the study constructs and theoretical frameworks employed in this thesis.

2.1 Math Anxiety

Math anxiety (MA) is a psychological construct characterised by feelings of tension, apprehension, or fear when dealing with mathematical activities, whether in academic situations or in everyday life (Ashcraft, 2002; Maloney & Beilock, 2012). The prevalence of math anxiety varies across populations. For example, Ramirez et al. (2018) reported that approximately 25% of college students in the United States experience math anxiety. My analysis of data from the 2022 PISA questionnaire (OECD, 2023), administered to 15-year-old students, showed that 25% of Swedish students and 20% of Finnish students reported experiencing math anxiety. These figures align closely with those reported in the United States (Ramirez et al., 2018). Inherent deficits in math achievement may be the key contributor to the development of math anxiety. However, while deficits in math achievement are certainly significant, there are also other factors to consider. To begin with, results from a twin-sibling study indicated that approximately 40% of the variation in math anxiety can be attributed to genetic predispositions (Wang et al., 2014). The remaining variance could then be explained by socio-environmental factors. Social and environmental causes can vary widely and are best understood as interacting factors. For example, studies have shown that parental support can alleviate math anxiety (Vukovic et al., 2013b), although, paradoxically, a math-anxious parent can pass on math anxiety through their frequent support (Maloney et al., 2015). Female students appear to internalise, to some extent, stereotypes suggesting girls have lower abilities in the STEM[2] subjects, which unfortunately can affect both learning processes and math anxiety (see review by Luttenberger et al., 2018).

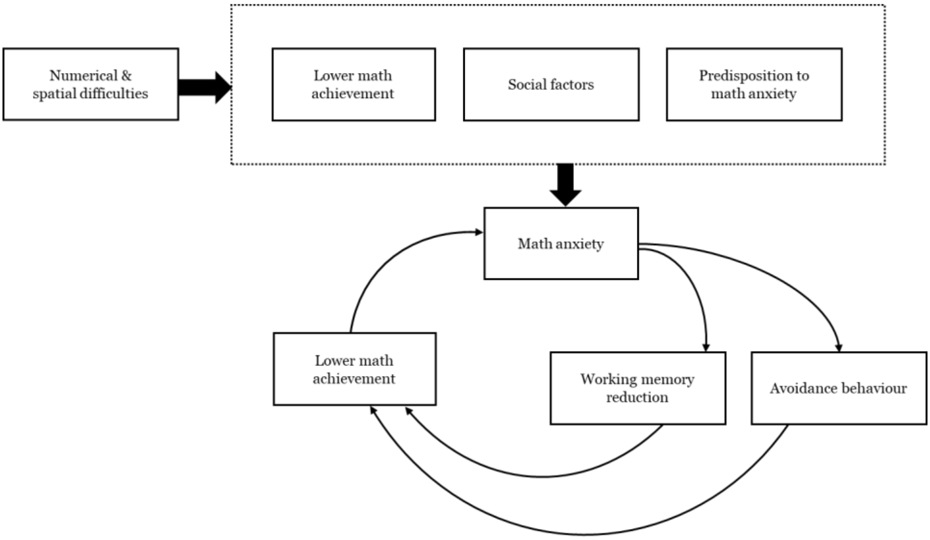

Figure 1. Plausible model of how math anxiety can emerge (Maloney, 2016).

We can assume that the emergence of math anxiety is not explained by deficits in math achievement alone. A plausible model for the emergence of math anxiety is provided in Maloney's work (2016; see Figure 1), which accounts for the aforementioned findings and illustrates how these developmental processes result in a cyclical relationship with math achievement. From a developmental perspective, data indicate that math anxiety peaks around Grade 9 and plateaus thereafter (Hembree, 1990). This plateau might be linked to the introduction of study choices. Avoidance behaviour could play a role, as students may steer clear of math-heavy subjects. However, Mononen et al. (2022) found that although students in grades 1 and 2 displayed low levels of anxiety, math anxiety was present and showed a significant negative correlation with arithmetic test scores, both cross-sectionally and longitudinally, but not with number processing tasks. This might suggest that facing numbers in non-cognitively demanding contexts does not trigger anxious responses in early childhood, whereas more cognitively demanding processes involving numerical calculations could instigate anxious feelings. Importantly, for older participants, the mere presence of numbers might trigger anxious states. This is suggested by Maloney and colleagues'

[2] STEM, abbreviation for science, technology, engineering and mathematics

(2010) experiments, where high math-anxious adults made significant errors in basic visual enumeration tasks compared to their non-anxious counterparts.

Measuring math anxiety

Specific anxieties have been studied for their associations with academic achievement as early as the 1950s. For example, Sarason (1957) found that general anxiety was not correlated with the sample's SAT (Scholastic Aptitude Test) or MAT (Mathematical Aptitude Test) scores, whereas a questionnaire measuring a more specific test anxiety construct was significantly correlated with both the SAT and MAT scores. The impact of specific anxieties on academic performance has gained greater recognition. As Richardson and Suinn (1972) recognised the need for a reliable measure of math anxiety both for research purposes and as a diagnostic tool for college populations, they developed the MARS (Math Anxiety Rating Scale). This 98-item questionnaire emerged as the first systematic instrument for measuring math anxiety (Ashcraft, 2002). Since the introduction of the MARS, several other instruments have been developed to meet different needs and populations. Suinn et al. (1988) created the MARS-E (MARS–Elementary), consisting of 26 items with a target population of younger school students, while Wigfield and Meece (1988) published the *Math Anxiety Questionnaire* (MAQ) for secondary school students. Both studies advocated for viewing math anxiety as a multidimensional construct. Suinn et al. (1988) identified factors such as math test anxiety and evaluation anxiety, while Wigfield and Meece proposed both a worry and an affective component of math anxiety. Other notable, widespread measures include the AMAS (The Abbreviated Math Anxiety Scale), introduced as a concise measure consisting of only nine items by Hopko et al. (2003), and the modified AMAS (Carey et al., 2017), which adapted the items from the AMAS to suit a younger English-speaking population (8-13-year-olds). More recently, Henschel and Roick (2017) published a questionnaire measuring a multidimensional math anxiety construct. Their findings support validity claims of a model consisting of cognitive and affective factors of math anxiety. The study by Henschel and Roick (2017) laid the groundwork for the development of the Swedish MARS-E, which is further discussed in the Methods chapter.

Most studies on math anxiety use self-report questionnaires. Since math anxiety is a hypothesized psychological construct, it can only be measured indirectly through indicators believed to represent the target construct. Self-report items about the participants' emotions are not the only possible type of indicators. In some instances, self-report questionnaires may even be considered a limitation, as they require: (I) self-inspection, which can be particularly challenging for children; (II) adequate literacy skills, which, again, may present difficulties for some children; and (III) responses that are not influenced by social desirability bias. Alternatively, math anxiety can also be measured using neuroscientific methods, such as magnetic resonance imaging or electroencephalography (Cipora et al., 2019), or through physiological indicators, such as skin conductance (Eidlin Levy & Rubinsten, 2021). Although these methods provide more objective measurements, they demand significant financial resources and raise other types of validity concerns. Specifically, it is uncertain whether variations in these measurements can be solely attributed to math anxiety or whether they might reflect other confounding variables.

State-trait math anxiety

Orbach et al. (2019) highlighted the considerable variability in both the instruments used to measure math anxiety and the results reported in empirical studies (see Table 1 for example). This variability may stem from the lack of universal diagnostic criteria for math anxiety. While the studies included in this thesis generally treat math anxiety as a unidimensional construct, we must still acknowledge alternative conceptualisations–specifically, those defined as multidimensional constructs.

To begin, we consider the two main dimensions of anxiety, namely trait anxiety and state anxiety. Trait anxiety refers to individual differences in the general tendency to respond to situations with consistent patterns of anxiety. Trait anxiety is relatively stable over time and across contexts. On the other hand, state anxiety is a temporary emotional response triggered by a specific situation. It reflects the interaction between an individual's trait anxiety and the characteristics of the situation they are facing. The degree of state anxiety a person experiences is influenced by how strongly the situation aligns with their underlying trait anxiety. A stronger match is more likely to trigger higher state anxiety (Endler & Kocovski, 2001).

Orbach et al. (2019) discuss the application of trait and state anxiety in the context of math anxiety. State math anxiety is a short-term reaction that arises in specific situations and increases autonomic nervous system arousal. Trait math anxiety, on the other hand, is more of a long-lasting personal tendency to see many math-related situations as stressful or threatening. Individuals with high trait math anxiety are more likely to experience state anxiety in response to math tasks. Over time, frequent experiences of state anxiety may influence more stable personality traits, such as trait anxiety (see Figure 2).

An important component of the trait-state anxiety relationship is *cognitive appraisal*. Lazarus (2001) described appraisal theory as "constantly evaluating relationships with the environment with respect to their implications for personal well-being" (p. 41). Appraisal concerns not only the event itself, but how it is interpreted. This interpretation can in turn shape how we feel. For example, if two individuals are about to take the same math test, they still might interpret the situation differently. Person A might think, “This test seems difficult—I’ll probably fail,” appraising it as a threat. Person B, on the other hand, might think, “This test seems challenging—I’m going to do my best,” appraising it as an opportunity to succeed. They are facing the same event, but interpreting it differently, resulting in different emotional responses. These interpretations reflect both the primary appraisal, what type of situation they believe they are facing, and the secondary appraisal, how they believe they can cope with it (Lazarus, 2001).

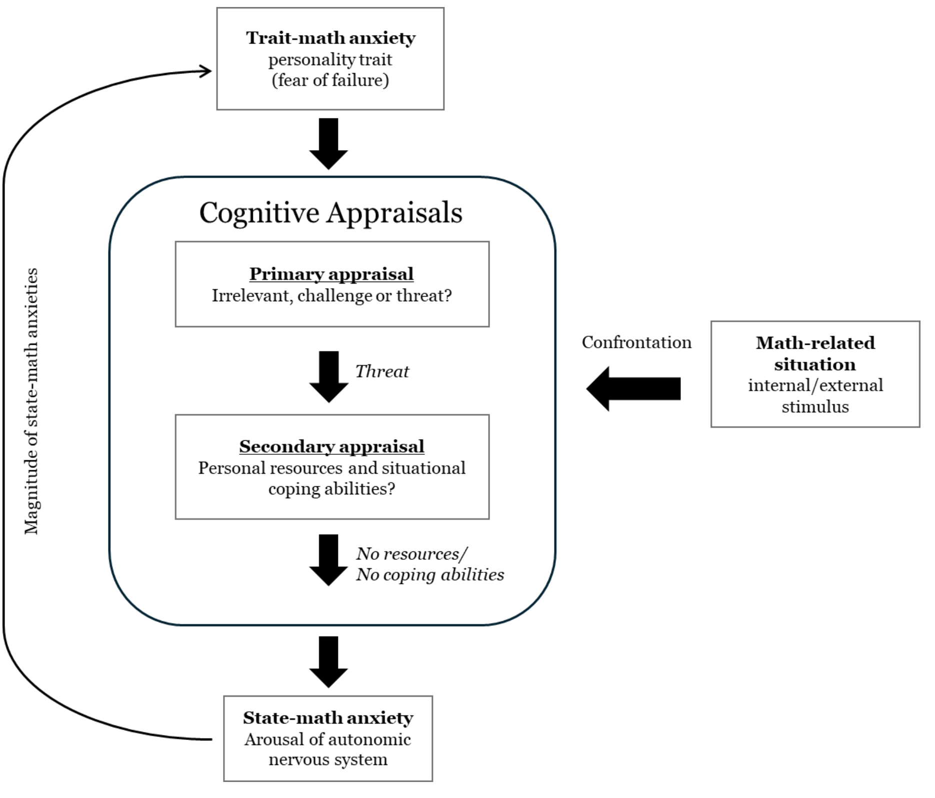

Figure 2. Orbach et al.'s (2019) model of state-trait math anxiety

Dimensionality of math anxiety

Orbach's (2019) application of the appraisal model (Figure 2) in a math context provides a plausible explanation of the dynamics of trait and state math anxiety. A highly math-anxious individual (trait) is more likely to appraise a math situation as threatening, and without a belief in their ability to cope with it, the situation triggers a state of anxiety. A high frequency of experiencing state anxiety can over time affect one's general tendency to perceive a situation as threatening—that is, their trait anxiety.

As mentioned earlier, math anxiety has been conceptualised as a multidimensional construct, in which one widely accepted framework separates the cognitive (worry) and the affective component of math anxiety. The cognitive component reflects negative expectations and fear of failure when dealing with mathematical problems, and it is characterised by preoccupation with worrisome thoughts. The affective component primarily affects the emotional and physiological responses associated with math, such as nervousness, tension, and other stress-related reactions (Henschel & Roick, 2017; Liebert & Morris, 1967; Wigfield & Meece, 1988). The distinction between these components is well supported in the literature. However, findings regarding how each component relates to other psychological or performance-based constructs remain somewhat inconsistent. For instance, Liebert and Morris (1967), Wigfield and Meece (1988), and Ho et al. (2000) found the affective component to be a considerably stronger predictor of math achievement compared to the cognitive component. Conversely, Henschel and Roick (2017) found that math achievement was more strongly related to the cognitive component compared to the affective component of math anxiety.

Another noteworthy conceptualisation of math anxiety's dimensionality distinguishes between math evaluation anxiety and math learning anxiety (Hopko et al., 2003; Plake & Parker, 1982). The evaluation component refers to testing situations (for example "Thinking about an upcoming math test one day before"), while the learning component relates to everyday classroom activities (for example "Starting a new chapter in a math book"). In some frameworks, these can be considered to be subdimensions within the broader cognitive and affective component (Henschel & Roick, 2020).

While the cognitive–affective distinction offers valuable insight into the complexity of math anxiety, treating it as a unidimensional construct remains a theoretically and practically sound approach, and has been widely adopted in previous research (Henschel & Roick, 2017).

2.2 Working Memory

Working memory is a central construct in cognitive psychology, crucial for a wide range of higher-order cognitive tasks, such as listening and reading comprehension, writing, following oral and spatial instructions, complex task learning, and more (Engle, 2010). Although the construct has received considerable attention, the definition of working memory varies across studies (Shah & Miyake, 1999). Baddeley and Hitch (1974) popularised the concept of working memory with their article "Working Memory", which proposed a model of working memory consisting of three systems: the central executive, the phonological loop, and the visuospatial sketchpad. The central executive is a controlling system, monitoring attentional processes. It is essential for directing flow and manipulation of information. The visuospatial sketchpad stores visual and spatial information, and is crucial for tasks involving spatial judgement, such as geometry-related tasks and navigation. The phonological loop is a two-part structure, which can store verbal information for a brief period and employs an articulatory rehearsal process, which can be compared to silent speech. This process effectively re-articulates and stores the information before it is lost (Baddeley, 2003). The phonological loop has been proposed to play a significant role in language development, specifically for learning new phonological forms of words (Baddeley et al., 1998). This conceptualisation of working memory provides a theoretical framework which could explain a cognitive system that "went beyond simple storage" (Baddeley, 2010, p. 138) allowing for a detailed, multifaceted modelling.

Prior to Baddeley and Hitch's (1974) seminal study, numerous articles were devoted to the topic of short-term memory. Baddeley and Hitch (1974) demonstrated that there was more to memory than just a short-term component. In their experiments, even when participants' digit span (a measure of short-term memory) was fully engaged, a substantial portion of their cognitive capacity remained available. This indicated the existence of a more complex system—working memory. Although these components were separate, the central executive within the working memory system was presumed to be active in both the input and the recall stages of information from the phonological loop (at the time referred to as phonemic buffer). The authors concluded that working memory was a system composed of multiple components, evidenced by empirical experiments (demonstrating the central executive and phonological loop), and theoretical reasoning (which suggested a third system for storing visuospatial memory should be included in this model). However, a limitation of the original model of working memory lies in the assumption of the central executive being mostly an attentional system, not providing a sufficient explanation for the interaction with long-term memory. This was highlighted by a case study of a patient with amnesia, who despite his condition, could recall extensive information immediately (Baddeley, 2003). This led to the introduction of a fourth main component to the working memory model: the episodic buffer. This component addresses some of the limitations of the initial model by facilitating interactions with long-term memory and integrating information from the two subsystems (Baddeley, 2000). Figure 3 depicts this revised construct of working memory.

Although the Baddeley and Hitch model emphasises the separate components of this cognitive system, these distinctions are not always clearly observable in practice. Engle et al. (1999) presented an alternative view in which working memory is treated as a modality-independent model. The two frameworks are not necessarily contradictory, as the former emphasises functional specialisation, while the latter promotes a more unitary concept that only separates the central executive from a general short-term memory span. The model used in this thesis is closely aligned with both frameworks, since the theoretical foundation is based on the Baddeley and Hitch model, while the statistical modelling has some resemblance to Engle's unitary construct.

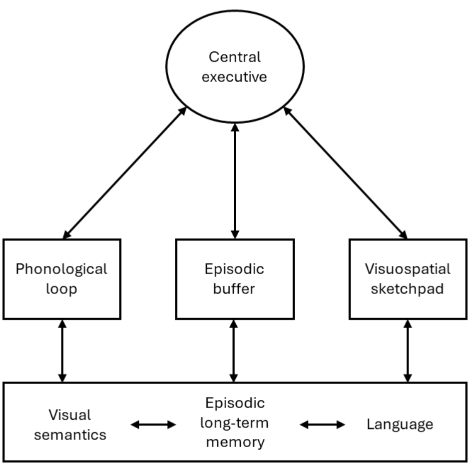

Figure 3. Working memory model as depicted by Baddeley (2002).

In educational contexts, empirical studies have consistently demonstrated a positive association between working memory and mathematics, both in cross-sectional studies (Swanson & Kim, 2007; Wiklund-Hörnqvist et al., 2016), and in longitudinal studies (De Smedt et al., 2009). This relationship holds true even when controlling for intelligence. In Alloway and Alloway's (2010) longitudinal study, five-year-old children were assessed on their working memory components and their general ability using an intelligence test. Six years later, the same children were administered reading and mathematical reasoning tests. Hierarchical regression models revealed that working memory accounted for more variance in the outcome variables than intelligence. While there have been mixed findings regarding which working memory components have the most significant impact on math performance, a meta-analysis found significant relationships from all components, with only minor variations (Friso-van den Bos et al., 2013). These findings underscore the importance of working memory in academic achievement, including mathematics.

2.3 Math Self-Concept

Self-awareness begins to develop in infancy and continues to evolve throughout childhood. A major milestone in this development occurs at around 18 months of age when many children, for the first time, demonstrate the ability to recognise themselves in a mirror as distinct from others (Amsterdam, 1972). Although self-awareness stabilises over time, the concept of oneself can still, and does still, change as interpersonal processes take place (Baumeister, 2010). *Self* is a complex construct. In short it can reflect on itself, gaining knowledge about itself, allowing itself to evolve through interplay with new input. From the concept of self, social psychology has focused on different aspects of self, resulting in a long list of varying terminology, such as self-concept, self-efficacy, self-esteem, self-perception, self-awareness and self-appraisal (Baumeister, 2010).

Self-concept and self-efficacy are two frequently used constructs in educational psychology. While related, they are sometimes difficult to distinguish. However, they can be differentiated based on specific conceptual dimensions. Self-concept is more about one's knowledge and perceptions of oneself, reflecting perceived competence. Whereas self-efficacy focuses on one's confidence in successfully carrying out a task, emphasising perceived confidence (Bong & Skaalvik, 2003). Despite their significant similarities, the studies in this thesis focus exclusively on self-concept.

Self-concept is a person's perception of themselves or the level of some property they have, and it is formed through their experiences with their environment and important people around them. A higher self-concept is considered important across various branches of psychology, including educational, personality, and sports psychology (Marsh & Martin, 2011).

It is a multidimensional construct with a hierarchical structure. At the highest level, the model includes a general self-concept, which then branches out into various academic self-concepts specific to school subjects, and non-academic self-concepts, such as social or physical self-concept. This particular definition and structure of self-concept is presented in Figure 4, which reproduces the model proposed by Shavelson et al. (1976).

Academic self-concept is shaped not only by of one's academic achievement but also influenced by the surrounding environment. Peer competence has been identified as an influential factor. Marsh and Park (1984) found that students in lower-ability schools generally had higher self-concepts compared to students in high-ability schools. This effect, known as the Big-Fish-Little-Pond Effect, describes how high-achievers in high-achieving environments may experience a negative impact on their self-concept. This phenomenon was also supported by a meta-analysis, which synthesised multilevel regression coefficients of classroom/school-level average achievement on students' academic self-concept, resulting in a small but significant negative effect (Fang et al., 2018).

Following the structure of self-concept depicted in Figure 4, we find more domain-specific areas as we move further down the hierarchy. The lower-level dimensions become more susceptible to volatility. Notably, these domain-specific self-concepts do not necessarily correlate with each other, even within academic domains such as math and verbal self-concepts[3]. While both math and verbal self-concepts correlate with achievements in their respective domains, and these achievements are themselves strongly correlated, the self-concepts do not correlate with each other (Marsh, 1986). This discrepancy is reflected in a study where researchers found non-significant effects between verbal self-concept and math achievement, despite the sample demonstrating significant moderate-to-strong effects between math achievement and math self-concept (Mueller & Winsor, 2016). Further support comes from a study demonstrating exceptionally strong correlations between grades and self-concepts within matching domains (math, verbal, and geography) compared to small correlations across domains (Podlogar et al., 2024).

[3] Verbal self-concept is sometimes quite broadly defined, though often it aims at describing one's perception of one's competence in written and reading abilities.

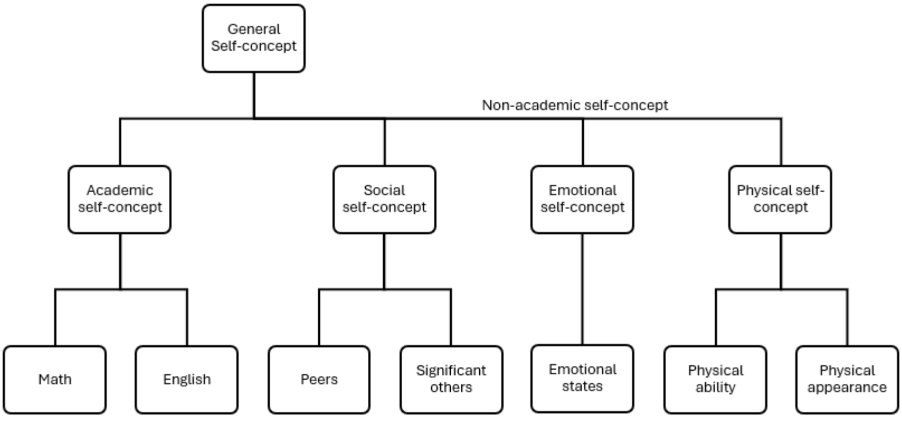

Figure 4. Simplified diagram of Shavelson et al.'s (1976) hierarchical model of self-concept.

From this description of the structure of self-concept and its multidimensionality, it becomes easier to appreciate the independence of domain-specific self-concepts, such as those related to math. This independence enables researchers to examine how math self-concept both influences and is influenced by math performance. There have been numerous studies examining the relationship between math self-concept and math achievement. The link between math achievement and math self-concept is likely reciprocal, as longitudinal cross-lagged panel models have shown that each variable explains unique variance in the other over time (Marsh & Craven, 2006; Marsh et al., 2005). How the relationship between academic self-concept and achievement is moderated by gender remains unclear (Wang & Yu, 2023), although mean differences favouring boys have been observed in both academic and math-specific self-concepts (Marsh & Craven, 2006; Skaalvik & Skaalvik, 2004).

Regarding the math anxiety – math performance link, anxiety may influence performance through motivational pathways. Students experiencing math anxiety often avoid math-related tasks (Lau et al., 2024) which over time can negatively impact their expectations regarding their math abilities. According to Expectancy-Value Theory, individuals' choices, persistence, and performance are partially explained by their beliefs in their own competence within a given activity. Achievement is influenced by expectancies for success and the value placed on the task. Both expectancies[4] and values[5] are thought to be influenced by ability

[4]Example item of expectancy: "How well do you expect to do in math this year? (not at all well - very well)" (Wigfield & Eccles, 2000, p. 70).
[5] Example item of usefulness: "Compared to most of your other activities, how useful is what you learn in math? (not at all useful - very useful)" (Wigfield & Eccles, 2000, p. 70).

beliefs (Wigfield & Eccles, 2000). Within this framework, individuals are more likely to engage in and persist with activities when they expect to succeed and/or perceive the task as valuable. In the context of math anxiety, students with low expectancy for success and low perceived value of math are likely to disengage from the activity, while those who view math as valuable may increase their effort despite their anxiety. Study IV examines multiple mediational processes by employing both working memory and self-concept in math as mediators in the math anxiety – math performance relationship.

2.4 Math Anxiety–Math Performance Relationship

All four studies included in this thesis address math performance, either as a means of assessing the validity of the math anxiety instrument or by exploring the relationship between math anxiety, performance, and potential mediating variables. Specifically, two measures of math performance were examined in Studies I, II, and IV: arithmetic fluency and number processing. Central to arithmetic fluency is the automatised retrieval of facts during single-digit arithmetic calculations (Balhinez & Shaul, 2019). Number processing, which relies even more heavily on automatic processes, is a fundamental cognitive skill that underpins a wide range of mathematical abilities (Geary, 2013; Träff et al., 2025), and can be used to screen for mathematical learning disabilities (Decarli et al., 2023; Hellstrand et al., 2024). This capability to map magnitudes has been shown to emerge well before verbal counting takes place. Studies have shown children of various ages can discriminate between two sets of number representations with a ratio of 2:3, and that even 6-month-old infants can distinguish between numerical representations with a ratio of 1:2 (Xu & Spelke, 2000).

The relationship between math anxiety and math performance has been extensively studied over the years. Various meta-analyses have consistently demonstrated a robust negative correlation between math anxiety and math performance. For example, Hembree (1990) reported a correlation of $r = -0.30$ for high school students. More recently, Namkung et al. (2019) found similar effect sizes with a mean $r = -0.27$ for primary school students, and $r = -0.36$ for secondary school students. Although these studies have primarily used measures of arithmetic computation, algebra, geometry, or other basic math operationalisations, there is indeed plenty of evidence suggesting a negatively directed association between math anxiety and number processing (Lindskog et al., 2017; Maldonado Moscoso et al., 2022; Maloney et al., 2011). While the association between these two variables is clear, the causality and direction of the relationship remain less certain.

Two competing theories have been proposed to explain the causal direction of the relationship between math anxiety and math performance. According to the Deficit Theory, deficits in math performance can lead to increased math anxiety, potentially through negative experiences with math tasks. In contrast, the Debilitating Anxiety Model proposes that anxiety itself affects the processing and retrieval of information, thereby reducing math performance. Together, these two contrasting theories contribute to a third, reciprocal model, in which math anxiety and performance influence each other bidirectionally (Carey et al., 2016).

Longitudinal studies are key in determining which comes first "the chicken or the egg", or in this case, anxiety or performance. Table 1 provides a summary of the studies that have investigated the relationship between math anxiety and math performance in longitudinal settings. As these studies employ varied designs, we cannot make direct comparisons, particularly because some do not test the bidirectional relationship. Nevertheless, Table 1 confirms that the evidence of the relationship in school settings remains inconsistent. This discrepancy might stem from a lack of universal diagnostic criteria leading to diverse operationalisations (Orbach et al., 2019), indicating that the measurement aspect requires more attention. Other contributing factors may include differences in math anxiety across various populations and the influence of varying environmental factors on the development of math anxiety.

A systematic review of the longitudinal effects can be found in the paper by Namkung with colleagues (2019), in which the authors extracted effect sizes for their meta-analysis. They coded separate longitudinal effect sizes from models where math anxiety predicts achievement, and vice versa. The results showed a correlation of -0.37 in both directions, supporting a bidirectional model. The overview of longitudinal studies presented in Table 1 shows a tendency to favour the deficit model. Notably, most effect sizes in Namkung et al.'s study were derived from samples that were older in age compared to those in Table 1, and the inclusion of various constraints and control variables in their analysis makes the results difficult to interpret.

Table 1. Longitudinal studies examining the math anxiety–math performance relationship in school settings

Study	Sample	Country	Grade Level	Design	Model	Controls	Math anxiety	Mathematics
Cargnelutti et al. (2017)	198	Italy	Primary G2 to G3	1 year; 2 waves	non-significant	Autoregressors, general anxiety	SEMA	Number skills
Ching (2017)	246	China	Primary G2 - G3	1 year; 2 waves	Debilitating	Number skills, working memory, test anxiety, intelligence, general anxiety	12-item test	Arithmetic fluency
Du et al. (2021)	2,789	China	Primary G4 - G6	2 years; 2 waves	Reciprocal	Autoregressors, math interest, math self-efficacy	5-item test	Multiple-choice 34 items
Gunderson et al. (2018)	634	USA	Primary G1 and G2	1 year; 2 waves	Reciprocal/ Deficit dominance	Autoregressors, motivation variables	CMAQ-R	Applied Problems
Guzmán et al. (2021)	451	Chile	Primary K - G2	2 years; 2 waves	Debilitating	No; bivariate correlation	CMAQ-R	Basic numerical skill
Kyttälä and Björn (2022)	116	Finland	Secondary G8 to G9	2 years; 3 waves	non-significant	Outcome expectancy and value	MAQ	Basic math skills
Ma and Xu (2004)	3,116	USA	Secondary G7 to G12	6 years; 6 waves	Reciprocal/ Deficit dominance	Autoregressors	2-item test	General math abilities
Pekrun et al. (2017)	3,425	Germany	Secondary G5 to G9	4 years; 5 waves	Reciprocal	Negative- and positive affect factors, intelligence, SES, gender	AEQ-M	School grades
Pellizzoni et al. (2022)	146	Italy	Primary G3 - G4	1 year; 2 waves	Debilitating	Working memory	AMAS	General math abilities
Sorvo et al. (2019)	976	Finland	Primary G2 to G5	1 year; 2 waves	Deficit	Autoregressors	MAQ	Arithmetic fluency
Sorvo et al. (2022)	848	Finland	Secondary G6 - G7	2 years; 2 waves	Debilitating	Autoregressors	AEQ	Arithmetic fluency

St Omer and Chen (2023)	335	Taiwan	Secondary G7 -G8	2 years; 4 waves	non-significant	Autoregressors, math cost	AEQ	School grades
Szczygieł et al. (2024)	348	Poland	Primary G1 - G2	1 year; 3 waves	Reciprocal	Autoregressors, working memory, intelligence, general anxiety	AMAS	Basic math skills
Vukovic et al. (2013a)	113	USA	Primary G2 - G3	1 year; 2 waves	non-significant	Autoregressors, working memory, reading achievement	12-item test	General math abilities
Wang et al. (2020)	444	Italy	Secondary age 14 - 21	1 year; 2 waves	Deficit	Autoregressors, intelligence, working memory	AMAS	School grades
Zhang et al. (2023)	601	China	Secondary G7 to G10	1 year; 3 waves	Deficit	Autoregressors, self-concept	4-item test	School grades

G = Grade, K = Kindergarten. SEMA = Scale for Early Mathematics Anxiety, AMAS = Abbreviated Math Anxiety Scale, MAQ = Math Anxiety Questionnaire, AEQ(-M) = Achievement Emotions Questionnaire(–Mathematics), CMAQ-R = Child Math Anxiety Questionnaire–Revised,

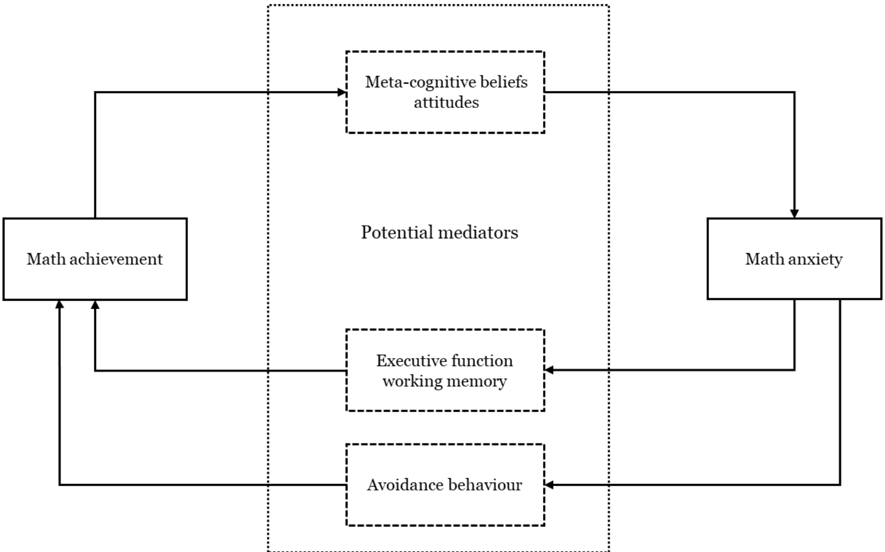

Figure 5. The math achievement – anxiety link described by Lau et al. (2024)

When examining longitudinal studies focusing on younger participants, math performance has been shown to exert stronger effects on future math anxiety (Gunderson et al., 2018; Ma & Xu, 2004; Sorvo et al., 2019). However, the difficulty in identifying anxiety effects on performance may arise from the high temporal stability of math performance, when modelled as an autoregressive variable (Pekrun et al., 2017). From these findings, I recognise that the driving variable and the onset of the bidirectional link likely lie in the math performance variable (which aligns with Maloney's model in Figure 1). Nonetheless, there is certainly evidence of variance to be explained from the emergence of math anxiety and its impact on performance. Although math performance appears to be the stronger predictor in the relationship (in younger participants), the present thesis adopts the perspective of the Debilitating Anxiety Model, which posits that math anxiety explains future math performance.

An important theoretical contribution to the math anxiety–math performance link, particularly relevant to study IV, is the recent review by Lau et al. (2024). In their model, the anxiety–performance link is mediated by self-beliefs and attitudes, as well as cognitive factors, including working memory (see Figure 5). Additionally, a third category in their model includes avoidance behaviours, which are related to long-term academic consequences. For example, Hembree (1990) identified a correlation of -0.31 between math anxiety and avoidance of high school math. However, avoidance behaviours are diverse and not always easy to capture. Lau et al. (2024) deconstructed these behaviours and created a conceptual representation that organises specific math avoidance behaviours along a timeline spectrum. At the "short-timescale" end of the spectrum, procrastination and loss of concentration were noted. Over time, math anxiety solidifies, and at the opposite end of the spectrum behaviours include reduced class engagement, avoidance of math courses, and even steering away from STEM careers. Avoiding math courses is certainly not favourable for one's math learning, but even smaller instances of avoidance behaviour could accumulate to a significant loss of learning opportunities. Furthermore, avoidance behaviour is characterised not only by the quantity of practice but also by its quality. Experiments have shown that students with high math anxiety exhibit impaired procedural math learning, suggesting that qualitative practice avoidance also occurs (Fioriti et al., 2025).

Although the model proposed by Lau and colleagues (2024) identifies self-beliefs specifically as mediators in the pathway from performance to anxiety, there is also empirical evidence supporting self-beliefs as mediators in the opposite direction—from anxiety to performance (Asare et al., 2025; Justicia-Galiano et al., 2017; Živković et al., 2023).

2.5 Working Memory in the Math Anxiety–Performance Link

Eysenck's (1979) review and conceptualisation of anxiety's impact on learning and memory highlight how task-irrelevant worry consumes working memory resources, thereby impairing cognitive performance. He argued that anxious individuals engage in more task-irrelevant processing, such as self-evaluative thoughts, which compete with task-relevant information for limited working memory capacity. This cognitive interference can lead to poorer performance on tasks that rely heavily on working memory, as evidenced by inferior digit-span performance among anxious individuals compared to their less anxious counterparts. Although anxious individuals can often compensate with increased effort, this does not fully outweigh the reduction in processing effectiveness. Ultimately, anxiety can impair the encoding of new information, and if this information is acquired under moments of failure, the process of retrieval is worsened.

Building upon Eysenck's (1979) framework, Eysenck and Calvo (1992) introduced the Processing Efficiency Theory. This model refined the understanding of anxiety's cognitive impact by distinguishing between *performance effectiveness* (quality of task performance) and *processing efficiency* (the ratio of performance effectiveness to effort). According to this theory, anxiety reduces working memory capacity primarily by occupying the central executive, and to some degree, the phonological loop. While anxious individuals may increase their effort to maintain task performance, this compensation comes at a higher cognitive cost, leading to reduced processing efficiency.

The Attentional Control Theory, proposed by Eysenck et al. (2007), extends the Processing Efficiency Theory by addressing some of its limitations. The Attentional Control Theory specifies that anxiety not only depletes working memory resources but also disrupts the balance between two attentional systems: the goal-directed system, responsible for top-down focus, and the stimulus-driven system (bottom-up), which is reactive to salient environmental cues. Under conditions of anxiety, the stimulus-driven system tends to become dominant, making individuals more susceptible to distractions, especially those that are threat-related. This shift impairs key executive functions such as inhibition (suppressing irrelevant information) and shifting (flexibly adjusting attention). While increased compensatory effort can preserve performance outcomes, it comes at the cost of reduced processing efficiency, especially as task demands increase.

Study III identified eight published studies that investigated working memory as a mediator of the anxiety–performance relationship. Since the database search was conducted, additional studies examining this type of model have emerged, especially those involving younger participants. In Italian Grade 3–5 students, the relationship between math anxiety and math performance was found to be partially mediated, both concurrently and longitudinally, by visuo-spatial and verbal components of working memory (Doz et al., 2024; Pellizzoni et al., 2022). A cross-sectional study including samples from Grades 3, 5 and 7 found that math anxiety impacts the visuospatial working memory component to a greater degree than verbal working memory (Živković et al., 2023). Furthermore, a recent study by Pelegrina and colleagues (2024) found that among executive functions, only the component more closely tied to working memory, updating[6], mediated the relationship between math anxiety and math performance. In contrast, inhibition and switching showed no significant association with math anxiety. This could indicate that the short-term memory component plays a crucial role, supporting the view that working memory is one of the underlying mechanisms in the anxiety–performance link, and that attentional control alone may not sufficiently explain the relationship.

[6] *Updating* is "Information that is maintained in working memory can be replaced with other information by means of an 'updating' process" (Nyberg & Eriksson, 2016, p. 2)

2.6 Validity Considerations

Validity is a fundamental concept in educational and psychological measurement. Although there is consensus on its importance, there is an ongoing debate on what exactly validity encompasses, both from a lexical and conceptual point of view. While some advocate for limiting validity to the traditional scientific framework, others argue it should include ethical and social consequences in its definition (Newton & Baird, 2016).

Traditional views of validity distinguish three types of validity: content validity, criterion validity, and construct validity. *Content validity* refers to the extent to which test items adequately cover the construct that is measured. *Criterion-related* validity addresses how well a test score predicts or correlates with an external criterion. *Construct validity* involves how well the test can measure the theoretical construct it is intended to measure. A turning point for this fragmented view of validity came with Messick's ideas of a unified concept of validity. The emphasis on validity then shifted from the test itself and its psychometric properties, to also include the interpretation and use of test scores (Messick, 1989), as well as evaluation of the consequences of measurement (Messick, 1995).

Messick (1995) highlighted two primary threats to construct validity: *construct underrepresentation* and *construct irrelevant variance*. Construct underrepresentation occurs when the measurement fails to capture important parts of the construct. Construct-irrelevant variance involves capturing excess variance that is associated with some other construct. In addition to identifying these threats, Messick proposed six interrelated aspects of validity that contribute to his unified framework. He referred to this broader conception as *construct validity*, emphasising the process of gathering many types of evidence. These are: content, substantive, structural, generalisability, external, and consequential aspects. While Messick's comprehensive framework provided a valuable theoretical approach for conceptualising validity evidence, it was often criticised for its complexity and lack of explicit guidelines for practical application. This prompted researchers and practitioners to find clearer, more structured ways to operationalise validity evidence, one of which is the development of Kane's argument-based approach.

O'Leary and colleagues (2017) point out that although contemporary validity work invests most of its focus on psychometric properties, it should include the interpretations and use of test scores. From this perspective, a test can have psychometrically valid metrics but still yield invalid results if interpreted or used incorrectly (Kane, 2013). Kane proposed an argument-based approach for evaluating appropriateness of the test interpretations and use, which he refers to as the Interpretation/Use Argument (IUA). Cook et al. (2015) applied Kane's four central inferences that form the validity argument: *Scoring* – how well the test responses or observed performance are reflected in an observed score. *Generalisation* – how well observed scores can generalise to an estimated trait value. *Extrapolation* – how well a test performance translates to real-world performance. *Implication* (decision) – applying the test score for decision-making or other conclusions (these could include ethical considerations or social consequences).

One of the more influential frameworks for validity is presented by The *Standards for Educational and Psychological Testing* (hereafter referred to simply as *Standards*), a document developed collaboratively by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME). This document provides guidelines and criteria for the development, evaluation, and use of tests in educational and psychological contexts. According to the *Standards*, validity refers to "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests." (AERA et al., 2014, p. 11), and is considered the cornerstone of evaluating and developing tests. In other words, according to the *Standards*, validity applies not only to the psychometric properties of the test, but also to the interpretation of test scores for specific uses (cf. Kane 2013). Therefore, each intended use requires a new interpretation, and with it, a separate validation effort. Importantly, validity is treated as "a unitary concept" (AERA et al., 2014, p. 14; cf. Messick), meaning that different sources of evidence should not be treated as independent types of validity, but rather as different aspects of the accumulated body of validity evidence.

Standards list five main aspects of validity: *test content, response processes, internal structure, relations to other variables,* and *consequences*. Test content refers to how well the content of the test represents the construct. Response processes refer to the extent to which the thought processes or behaviours of test-takers align with those intended by the test developers. Internal structure involves evaluating statistical relationships among test items to determine whether the structure reflects the intended construct. Relations to other variables refers to how a test score relates to other measures in way that is theoretically expected, for example, through convergent and discriminant validity, or predictive and concurrent relationships. Consequences address the intended and unintended effects of test use. A major focus of consequences is typically on decision-making based on scores, as well as issues of fairness across distinct groups.

2.6.2 Validity Concerns in Self-Report Questionnaires

Standards (AERA et al., 2014) provide a comprehensive validity framework that is applicable to my research. However, since this thesis relies partly on questionnaire data, I will end this chapter by addressing some questionnaire-specific issues that could impact the validity of the findings.

Self-report questionnaires are widely used in the social sciences, particularly in educational and psychological research. Despite their widespread use, they are susceptible to various validity threats. Chan (2010) identifies four main problems commonly associated with self-report questionnaires.

1. *Construct Validity*: self-report data may suffer from construct-irrelevant variance as responses could include systematic errors stemming from confounding variables, such as language comprehension.
2. *Common method variance*: observed correlations between self-reported variables might be inflated because of shared method effects, rather than actual relationships between the constructs.
3. *Social desirability bias*: respondents may answer in ways they believe is socially favourable, resulting in skewed responses.
4. *Perceived inferiority of self-report data*: there is often a perception that self-report measures are less valid than non-self-report measures. However, this belief often disregards the fact that non-self-report data can also suffer from validity issues.

While these issues highlight important considerations, Chan (2010) cautions against overgeneralising the limitations of self-report methods. He emphasises that self-report data are not inherently flawed, and that their validity relies on careful design, administration, and validation efforts.

3 Methodology

3.1 Sample and Procedure

My PhD project is conducted primarily within the research project *Choking Under Pressure*, a longitudinal study that collected data from Grade 4 students over four subsequent semesters (See Figure 6 for the project design). Studies I, II, and IV use data collected as part of this project. Study III is a literature review that synthesised effect sizes from multiple samples to estimate an average effect. Studies I and IV used data from three measurement points collected within the Swedish *Choking Under Pressure* project. Study II combined data from one measurement point from *Choking Under Pressure* (Spring 2023) with corresponding data from the Finnish projects *Samsyn* and *iFeelMath*.

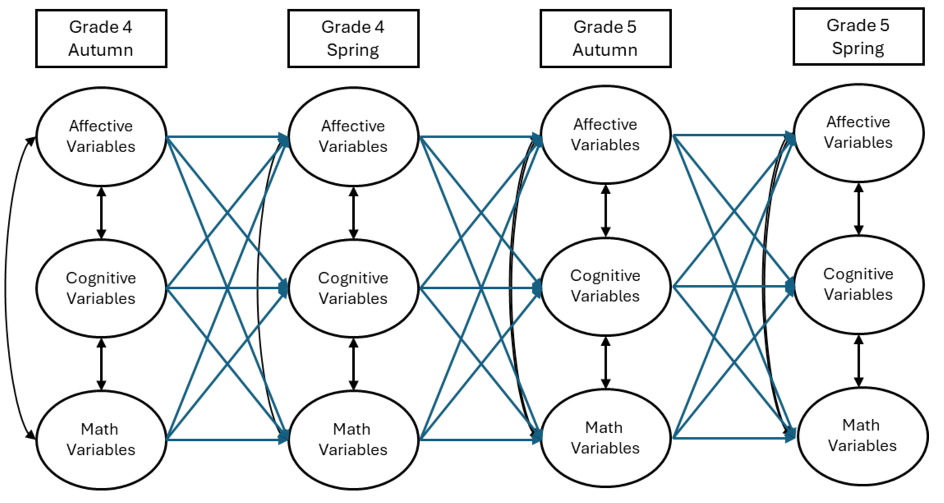

Figure 6. Longitudinal relationships of affective, cognitive and math performance variables measured in the project Choking Under Pressure.

Study III was conducted as a meta-analysis, in which relevant literature was identified using predetermined criteria selected to address the research questions. The total sample included 16,589 participants, in which 1,824 participants were used to answer the second research question. These participants were drawn from 66 unique samples, resulting in 150 effect sizes for the first question and 15 for the second. The data were extracted from 57 articles published between 1997 and 2020[7]. The included samples varied in age and origin, spanning across

[7] In study III, there were no constraints on publishing dates in the literature search-phase.

nearly 20 countries from four continents, and included participants at the primary, secondary and university levels. No constraints were placed on sample characteristics during the literature search.

Studies I and IV used data from the first three measurement points of the Choking Under Pressure project. Of the approximately 700 invited students, 436 provided consent. Study II used the second measurement from this Swedish sample (n = 395) and combined it with the corresponding data from the two Finnish projects: Samsyn (n = 295), and iFeelMath (n = 332). Across all three projects, gender distribution was approximately equal, resulting in a total sample size of 1,022 participants (52.6% girls).

Missing Data

Missing data at the individual level were not examined for the meta-analysis in Study III or for the Finnish projects in Study II. For Studies I and IV, missingness was assessed using Little's MCAR test (Missing Completely At Random; Little, 1988), which suggested that data were not missing completely at random (χ^2 = 1140, df = 966, p <0.001). Because test sessions were integrated into the standard school schedule, and participation was mandatory for everyone (even those without consent as they were filtered out later), the likelihood of students selectively avoiding test sessions was reduced. This reduced the plausibility of MNAR (Missing Not At Random) mechanisms and supports the assumption that the data were MAR (Missing At Random).

3.2 Assessments

Within the Swedish project *Choking Under Pressure* (Figure 6), we administered a math performance battery, a questionnaire, and a cognitive battery. The questionnaire included items assessing math anxiety, math self-concept, math interest, test anxiety, and demographic variables such as age, mother tongue, and gender. Math self-concept and interest were measured using items from Marsh et al.'s (1983) Self-Description Questionnaire (SDQ), with three items per construct. Responses were rated on a 5-point Likert scale. Response options ranged from “False” to “True”, with items such as “I am good at mathematics” for self-concept and “I am interested in mathematics” for the interest.

Test anxiety was measured with a Swedish adaptation of Gierl and Bisanz's (1995) School Test Anxiety Survey (STAS). The Swedish version consisted of six items on a 4-point Likert scale. Response options ranged from “Not at all nervous” to “Very nervous” and included items like “How nervous do you feel before taking an exam at school?”.

The cognitive battery employed in Study IV included tests of operation span (verbal and math), digit span, letter span, and spatial span. In the operation span verbal task, participants were asked to remember an increasing sequence of letters, with a true/false statement presented between each letter that they had to judge. In the operation span math task, participants also memorised letters but solved simple one-digit arithmetic problems between each letter. The spatial span task required participants to recall a visual pattern, including both location and sequence, on a 4×4 grid. In the letter span task, participants saw an increasing number of letters, which they needed to recall in the correct order. The digit span task followed the same procedure but used numbers instead of letters.

In the context of this thesis, math performance refers to the FUNA-DB (Functional Numeracy Assessment–Dyscalculia Battery), an online screener for dyscalculia. The battery consists of 6 tasks: (1) digit-dot match, (2) number comparison, (3) number series, (4) single-digit subtraction, (5) single-digit addition, and (6) multi-digit addition and subtraction. The first two tasks serve as indicators of a number processing factor, while the remaining four assess arithmetic fluency. Although the instrument is relatively new, there is already a vast amount of evidence supporting its validity (Hakkarainen et al., 2025; Hellstrand et al., 2024; Räsänen et al., 2021). Studies I, II, and IV assessed math anxiety using the Swedish MARS-E, which is further discussed in the next chapter.

3.3 The Swedish MARS-E

While math anxiety measures have been used with older Swedish-speaking students (Korhonen & Räsänen, 2018), to my knowledge, no formally validated instrument exists for younger Swedish-speaking students. Therefore, we saw the need for such a tool, particularly given the research within this thesis' focus on the relationships between math anxiety and other cognitive and affective constructs.

As described previously, math anxiety instruments have evolved considerably over time. In short, the MARS (Math Anxiety Rating Scale) was originally developed as a 98-item scale (Richardson & Suinn, 1972), and has since been adapted and shortened for different purposes and target groups (e.g., MARS-R, Plake & Parker, 1982; MARS-E, Suinn et al., 1988). Henschel and Roick (2017) introduced a shortened version of the Mathematics Anxiety Rating Scale–Elementary (MARS-E), consisting of 36 items, equally divided between cognitive (18 items) and affective (18 items) factors.

Given the age and diversity of the Swedish sample (4th Grade students), a shorter and more accessible instrument was needed to ensure that all participants could both comprehend the questions and maintain focus during the test sessions. Based on the groundwork laid by Henschel and Roick (2017), the Swedish MARS-E was translated and adapted to a 16-item version, through multiple pilot iterations. Eight items were retained from the cognitive factor and eight from the affective factor.

The cognitive factor is associated with worrisome thoughts and negative expectations related to failure in mathematical contexts. The affective factor is associated with feelings of nervousness of dealing with mathematical problems (Henschel & Roick, 2017). Research on test anxiety provides a similar distinction. Liebert and Morris (1967) described the worry component as a *cognitive expression of concern about one's own performance*, while emotionality (which is the affective component, Wigfield & Meece, 1988) was described as *autonomic physiological reactions*.

In the Swedish MARS-E, the items are phrased to assess either nervousness or worry related to mathematical activities. Responses are measured on a four-point Likert scale: "Not at all worried/nervous", "A little worried/nervous", "Quite worried/nervous", and "Very worried/nervous".

3.4 Study Methodology

The methods used in the four studies consist primarily of a latent variable approach, using Confirmatory Factor Analysis and Structural Equational Modelling in Studies I, II and IV, and meta-analysis in Study III.

3.4.1 Latent Variable Modelling

Studies I, II, and IV incorporate confirmatory factor analysis (CFA) and invariance testing to evaluate measurement models. CFA is a type of structural equation model that focuses exclusively on the measurement model and is well suited for instrument development. Study I emphasised the measurement model, as its primary objective was to assess the construct validity of the Swedish MARS-E. A CFA model must be specified by the researcher (Brown, 2015), a process that involves defining how observed variables (indicators) load onto specific latent factors. This process requires a theoretical foundation to determine how the underlying constructs are represented by measured indicators (Kline, 2016). In study I, the model structure was guided by Henschel and Roick's (2017) work. In Studies II and IV, model specification was based on a combination of findings from Study I and established theory concerning the math anxiety–math performance relationship (Eysenck et al., 2007; Justicia-Galiano et al., 2017; Lau et al., 2024).

Employing CFA in scale development has many strengths. Its primary focus is on the structure of the construct, allowing for a thorough examination of dimensionality. Both convergent and discriminant validity can be evaluated, while also accounting for measurement error(Brown, 2015). For each factor loading (a metric reflecting how strongly the item is related to the latent factor), CFA also estimates a residual variance for the corresponding item. This residual represents the portion of variance not explained by the latent factor, also referred to as unique variance. Unique variance consists of both systematic and random variance that is not shared with other items in the factor model.

Maximum likelihood (ML) estimation assumes normality in the endogenous variables, which typically requires continuous data. Although there is no strict rule for when a scale can be treated as continuous (Kline, 2016), some suggest that data with more than five response options may, in some cases, be treated as continuous (Liu et al., 2017). The scale for the MARS-E is considered ordinal with a maximum of four response options. Therefore, the Weighted Least Squares Mean and Variance adjusted (WLSMV) estimator was used, as it is specifically designed to handle the ordinal nature of categorical variables.

Study IV employed Structural Equation Modelling (SEM), which extends the measurement model by adding a structural component. SEM is well suited for testing theoretical causal relationships between latent variables (Hayduk et al., 2007) and it also accounts for measurement error.

3.4.2 Multiple-Group and Invariance Testing

Multiple-group CFA is a procedure in which measurement models are tested across different groups simultaneously, making it particularly useful for assessing group-specific correlations and the validity of a construct across populations. This approach helps address a key question: Does the scale measure the same construct across different groups? Invariance is an essential property of a test and a prerequisite for comparing means between groups. In the case of the Swedish MARS-E, establishing invariance is particularly important, as mean differences in math anxiety have been reported in previous studies (e.g., Hart & Ganley, 2019; Hembree, 1990).

Measurement invariance is a technique used to assess whether a construct is measured equivalently across groups or timepoints. Building upon multiple-group analysis, Studies I, II, and IV examined invariance across gender groups. The process involves gradually constraining parameters to equality across both groups and comparing the constrained model against the less constrained model. If a baseline model, typically referred to as the configural model, fits the data, the researcher can proceed with stepwise constraints. These include factor loadings (metric model), intercepts or thresholds (scalar model), and in some cases also the residual variances (strict model, used in Study I).

In longitudinal studies, measurement invariance is essential when the goal is to assess the development of math anxiety over time. Valid comparisons across time points require the test to remain invariant, meaning it must measure the same construct consistently over time. The process begins with specifying a longitudinal CFA. If the configural model fits, stepwise constraints are introduced across time points, starting with factor loadings (metric), then intercepts or thresholds (scalar), and finally residual variances (strict).

3.4.3 Mediation

Mediation analysis is a statistical method used to understand how an independent variable (x) influences a dependent variable (y) through a mediator (m), which may represent the mechanism driving the x–y relationship. In contrast, moderation tests how different conditions or subgroups affect the strength or direction of the x–y relationship. In the causal chain of mediation (see Figure 7), mediation effects can be claimed if the following conditions are met:

1. X is significantly related to m (*path a*),

2. M is significantly related to y (*path b*), and

3. When paths *a* and *b* are accounted for, the previously significant direct path (*c'*) between x and y becomes non-significant.

Because constructs in the social sciences are rarely influenced by a single factor, mediation is best viewed as a continuum. Partial mediation occurs when the direct path (*c'*) remains significant but is reduced in magnitude (Baron & Kenny, 1986).

Mediation analysis with one mediator, such as the model depicted in Figure 7, is referred to as *simple mediation*. The mediating effect, or indirect effect, is calculated as the product of path *a* and path *b* (indirect effect = *ab*). The total effect of x on y is then computed as the sum of the direct path (*c'*) and the indirect effect (*ab*). This approach assumes that at least the mediator (m) and outcome variable (y) are measured on a continuous scale.

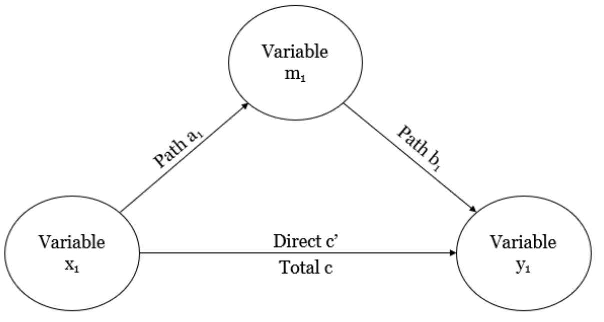

Figure 7. Conceptual model of the mediation chain.

Multiple mediation occurs when more than one mediator is included. For example, adding a second mediator (m_2) to the model in Figure 7 introduces new paths a_2 and b_2, with the indirect effect through m_2 calculated as a_2b_2. There are several advantages to using multiple mediation rather than conducting separate simple mediation analyses. First, it allows us to determine whether m_1 mediates the x–y link while controlling for m_2. Second, we can compare the magnitudes of the indirect effects, which can be particularly interesting when testing competing theories. However, as in multiple regression analysis, collinearity can become a concern when including multiple mediators in the same model. High correlations between mediators could compromise their unique contributions (Preacher & Hayes, 2008).

Preacher and Hayes (2008) discuss the assessment of indirect effects and how to determine their significance. Although a common approach is to use the Sobel test, bootstrapping is often more robust, as it does not assume a normal distribution of the test statistic. Bootstrapping involves resampling from the data with replacement to generate a distribution of indirect effects, which is then used to estimate confidence intervals. These confidence intervals are in turn used to determine the statistical significance of the indirect effects. In Study IV, we used multiple mediation to examine the mediating effects of both working memory and math self-concept on the math anxiety–math performance link. Bootstrapping with 5000 resamples was used to estimate confidence intervals. If the 95% confidence interval did not include zero, the effect was interpreted as statistically significant.

3.4.4 Meta-Analysis

In Study III, we addressed two research questions:

(I) What is the mean correlation between math anxiety and working memory?

(II) What is the indirect effect of math anxiety on math performance, with working memory as a mediator?

To answer these questions, we conducted a systematic literature search across 13 databases, screening approximately 1,900 abstracts and reviewing 97 full-text articles, of which 57 were included in the final analysis. Since the key outcomes of interest were correlation coefficients and indirect effects, meta-analysis was deemed the appropriate method for synthesising the findings.

Meta-analysis was formally introduced by Glass in the 1970's, where the author described it as:

> "Meta-analysis refers to the analysis of analyses. I use it to refer to the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings." (Glass, 1976, p. 3)

We applied meta-analysis techniques to synthesise the effect sizes extracted from the collected data. The primary aim of meta-analysis is to estimate a summary effect from a group of studies, whether treatment-based or correlational. As Borenstein et al. (2009, p. 6) describe, the summary effect is "nothing more than the weighted mean of the individual effects.". Two commonly used models in meta-analysis are the *fixed-effect model* and the *random-effects model*. The fixed-effect model assumes a single source of variation, namely sampling error, whereas the random-effects model assumes two sources of variation: sampling error and study variation[8], reflecting the possibility of multiple true effects, hence the plural form of "effects" (Borenstein et al., 2009).

Ordinary meta-analysis (both fixed- and random effects models) assume that effect sizes are independent (Cheung, 2014). In the case of Study III, this assumption would require that each effect size originates from a unique sample. However, this was not the case, as the number of effect sizes exceeded the number of samples, indicating non-independence. Several strategies are available to account for such dependency. One option is simply to retain one effect from each study/sample and discard

[8] The fixed-effect model would be suitable in e.g. analysing clinical trials using the same treatment, whereas the random-effects model would be suitable when studies vary in key characteristics like methods or populations.

the rest. However, this reduces statistical power and affects the precision of the estimated mean effect. Other more viable alternatives include the three-level meta-analysis model (Cheung, 2014) and the robust variance estimation technique (Hedges et al., 2010), both of which account for dependent effect sizes without discarding valuable data.

The three-level meta-analysis model assumes a hierarchical structure where level 1 represents participants, level 2 represents studies, and level 3 represents clusters. The cluster level could comprise cultural groups, research teams, or samples. In Study III, level 3 could refer to unique samples, as some studies reported multiple effect sizes from the same sample. One advantage of the three-level model over the robust variance estimation technique, is that it allows the estimation of heterogeneity at both level 2 and level 3. Like the three-level model, the robust variance estimation accounts for dependency structures, but unlike the three-level model, it does not estimate variance components across multiple levels. While the multilevel approach provides richer output, it also requires more complex statistical modelling. Since the primary goal was to address dependency in the data, and given the limited number of studies available for the second research question, robust variance estimation was adopted for the main analyses. However, as the robust variance estimation method is a form of meta-regression (Hedges et al., 2010), a random-effects model was used for the subgroup analysis, which involved categories with more than three conditions. While it would have been possible to dummy-code new categories and include them as predictors, the random-effects model was deemed more appropriate for this part of the analysis.

3.5 Ethical Considerations

Ethical considerations are fundamental in all research involving human participants, and especially when studying children. Although the studies included in this thesis did not involve any procedures that posed a risk to participants' physical or mental health, several other ethical aspects still needed to be addressed. The research process was outlined prior to carrying out the research project *Choking Under Pressure* and was submitted for ethical review by the Swedish Ethical Review Authority[9]. The Finnish projects were reviewed by Finnish ethical review boards. The following ethical aspects were considered in Studies I, II and IV.

Informed consent and *voluntary participation*. Given that the participants were primary school students, informed consent was obtained from their legal guardian(s). Parents received a brief (general)

[9] Reference number: 2020-05982

description of the project, as well as a detailed description of project's purpose, procedures, potential risk and data confidentiality. Children were also informed about the study in an age-appropriate manner. Tests were administered during regular class time, in which all students participated. The children were informed that only responses associated with consent were to be processed, and naturally participants were not forced to complete the tests if they did not wish to. For the Finnish projects in Study II, ethical approval was granted by the Board of Research Ethics at Åbo Akademi University, and consent was obtained from both parents and teachers.

All data were collected and stored in accordance with The General Data Protection Regulation (GDPR). Participants were informed they could withdraw at any time, as well as have their data deleted (or access their data) without providing a reason. Responses were anonymised using unique identification codes. No identifiable information was shared in publications or presentations. The only exception was that each student's teacher received stanine scores from the math tests. Only authorised researchers had access to the raw data, which was securely stored and handled in compliance with institutional guidelines.

Study III extracted data from published articles. Although we had no control over the original researcher's procedures, all included studies were published in peer-reviewed journals, which are expected to uphold established research ethics standards.

4 Summary of Studies

This chapter provides a summary of all four studies that are included in this thesis. All studies investigate math anxiety and its relationship to performance measures in mathematics. Study I assesses validity properties of the Swedish MARS-E. Study II combines data from three separate research projects, utilising samples of Swedish, Finland-Swedish, and Finnish students, and compares the math anxiety–math performance relationships across cultural groups. Study III is a meta-analysis study which synthesises correlations between math anxiety and working memory as well as indirect effects of math anxiety and math performance with working memory as a mediator. Study IV builds on the third study by examining the mediators of the math anxiety–math performance link using longitudinal data and latent variables.

Study I

For numerous students, mathematics is considered one of the most challenging subjects and is often associated with math anxiety starting from an early age. Math anxiety is characterised by feelings of fear, tension, and apprehension related to math tasks, which can negatively affect both emotional well-being and performance in mathematics. Research, especially at the secondary school level, consistently shows a significant negative correlation between math anxiety and math achievement. Approximately one in four college students in the United States report experiencing math anxiety. While gender differences in the levels of math anxiety are frequently observed, its effect on math performance is typically not moderated by gender.

In Study I, we assessed the newly developed instrument for measuring math anxiety in Swedish-speaking students, called the Swedish MARS-E. It was translated from the German MARS-E and revised to suit a Nordic context (e.g. one question asked about nervousness that arises from receiving change at the store, a scenario less relevant in Sweden). The original German MARS-E consisted of 36 items intended to measure two components of math anxiety: a cognitive and an affective component. The Swedish MARS-E reduced the original instrument from 36 to 16 items, retaining eight items from the cognitive factor and eight from the affective factor. This shortened version allowed for easier administration to a young and heterogeneous sample.

Reliability was assessed using the test-retest method, with measurements from three timepoints spread over three semesters (one per semester).

Internal consistency was also examined for each measurement. The results suggested that the Swedish MARS-E was consistent over time, as well as demonstrating high levels of agreement among the items.

The validity aspects presented in Study I can be divided into three categories: construct validity, criterion validity, and content validity. Content validity was addressed partly by the researchers behind the German version, and partly by the research team conducting Study I. As part of this process, the instrument was piloted and revised based on the results and feedback from the teachers and students who completed the test. This was an iterative process, with the instrument piloted a total of three times across two different schools.

Construct validity, which concerns how well the instrument measures the intended construct, was assessed through confirmatory factor analysis. Both longitudinal and gender invariance tests were carried out to evaluate the internal structure of the construct across groups and timepoints The dimensionality of the math anxiety construct was examined, and although a two-factor model—mirroring the German MARS-E—fit the data, the two factors were highly correlated, suggesting no meaningful distinction between cognitive and affective components. Given that a one-factor solution also fit the data, the principle of Occam's razor was applied, and the Swedish MARS-E was retained as a unidimensional model for ease of interpretation. Results from the invariance tests indicated that the instrument measured the construct consistently across both time and gender groups.

Finally, criterion validity was examined by correlating the math anxiety factor with other known criterion variables: test anxiety, math performance, self-concept in math, interest in math, and gender. These correlations were consistent with previous findings. Given the strong correlation between math anxiety and test anxiety, a confirmatory factor analysis was conducted to test a one-factor model including items from both scales. The model demonstrated poor model fit, suggesting that the two instruments measure distinct constructs. This finding provided evidence of discriminant validity. Additional support was found in the systematic differences in how math anxiety and test anxiety correlated with other criterion variables. Study I contributed a new, valid and reliable instrument for measuring math anxiety in Swedish-speaking primary school students.

Study II

Study II builds on Study I by further examining the Swedish MARS-E and its relations to math performance. This cross-sectional study tested the Swedish MARS-E in three samples: a Swedish, a Finland-Swedish, and a Finnish-speaking sample. All groups were assessed using the same instruments during the same semester (spring 2023) in Grade 4. The combined sample included 1022 fourth-grade students from Sweden ($N_{\text{Swedish-speaking}} = 395$) and Finland ($N_{\text{Finnish}} = 332$; $N_{\text{Finland-Swedish}} = 295$). The Swedish MARS-E was administered to both Swedish-speaking groups and was translated into Finnish for the Finnish-speaking group.

Different aspects of validity and reliability were assessed in each sample. Dimensionality analysis revealed distinct factor structures across the language groups: a one-factor model fitted the Swedish-speaking groups, while a two-factor model, consisting of a cognitive and affective math anxiety component, fitted the Finnish-speaking group best. Structural relations between math anxiety and both arithmetic fluency and number processing were examined within each sample, and a multiple-group analysis was conducted to assess potential gender differences.

Differences between math anxiety and different math performance measures (number processing and arithmetic fluency) were more apparent in the Finnish-speaking sample. Latent means of math anxiety, from a constrained confirmatory factor analysis, showed significant gender differences, with the largest discrepancies found in the Finland-Swedish sample.

In the structural models, path coefficients showed stronger associations between anxiety and arithmetic fluency compared to number processing, a pattern consistent across all samples. In the Finnish-speaking group, no significant paths were observed between the worry component and math performances. Multiple-group analyses indicated that the predictive paths for girls were weaker than those for boys when modelling number processing. For arithmetic fluency, gender differences were less consistent. A significant correlation between math anxiety and number processing was found for girls in the Swedish sample, but not in either of the two Finnish samples.

Study III

Previous research has consistently demonstrated a significant negative relationship between math anxiety and math performance. Recent research has applied Attentional Control Theory to the context of math anxiety. According to this framework, anxiety impairs or depletes the cognitive resources necessary for effective mathematical processing.

Study III aimed to extract all published, peer-reviewed, studies written in English on the topic of math anxiety, working memory and math performance. Two research questions were posed.

(1) What is the mean correlation between math anxiety and working memory?
(2) What is the mean indirect effect between math anxiety and math performance while accounting for working memory as a mediator?

A total of 1346 unique studies were identified through database searches using a standardised search string. All abstracts were screened, and based on predetermined criteria, 97 studies were considered eligible for full-text analysis. Of these, 57 studies were deemed suitable for answering the first research question, and 8 were deemed suitable for the second research question.

The meta-analysis revealed a small but significant negative correlation between math anxiety and working memory. Subgroup analyses showed that the strength of this relationship varied by age, with participants aged 13 and older demonstrating a stronger negative correlation. The strength of the relationship also varied depending on the type of working memory test, as numerically based tasks showed a stronger negative association with math anxiety, compared to non-numerical tasks.

Fifteen effect sizes were included in the analysis addressing the second research question concerning mediation. The analysis revealed a significant indirect effect between math anxiety and math performance, suggesting that working memory at least partially mediated this relationship. However, the small number of available studies limits the external validity of these findings. Therefore, the results should be interpreted as descriptive rather than prescriptive, and generalisations should be made with caution.

In sum, Study III contributes evidence that the published literature demonstrates a small but significant negative correlation between math anxiety and working memory, and highlights working memory as a potential mediator in the math anxiety–math performance relationship.

Study IV

The fourth study extends the work of Study III through the empirical application of its theoretical framework. While Study III reviewed models where working memory acts as a mediator in the math anxiety–math performance relationship, Study IV aimed to test this mediation using a different method. In contrast to the cross-sectional studies included in the meta-analysis, which often relied on observed sum scores, Study IV employed a latent-variable approach within a longitudinal design, using T1, T2, and T3 measurements.

Predictors and mediators were selected from the same time point, and outcome variables from the subsequent time point. This was done in two waves, using T1 and T2 variables for wave 1, and T2 and T3 variables for wave 2. Including two waves allowed us to evaluate the stability of the model's temporal pathways. An additional affective mediator, math self-concept, was introduced and contrasted with the cognitive mediator, working memory.

The results were analysed with structural equational modelling. Each mediator (working memory and self-concept in math) was introduced in a stepwise manner and tested with both arithmetic fluency and number processing as outcome variables. Significant mediation was evaluated using the bootstrap resampling technique, which estimated confidence intervals for the indirect effects of the mediators. The main model included both mediators and both outcome variables, along with a multiple-group analysis by gender. This allowed us to assess the direct effects of math anxiety on performance outcomes under two mediation conditions, while also estimating gender-specific effects.

The results extended the findings from Study III by confirming that working memory partially mediated the influence of math anxiety on both number processing and arithmetic fluency, across both longitudinal waves. Math self-concept also played a mediating role in the math anxiety–math performance relationship, although its indirect effect was less pronounced than that of working memory.

In conclusion, Study IV provided a more nuanced understanding of the math anxiety–math performance relationship by employing latent variables and two longitudinal waves to test the stability of the effects over time. The study confirmed that both working memory and math self-concept were important mediators, influencing both arithmetic fluency and number processing. Gender-specific analyses revealed minor variations across gender groups.

5 Discussion

The purpose of this thesis was two-fold. First, it aimed to develop and validate a math anxiety rating scale for Swedish-speaking primary school students in a Nordic context. Second, it sought to explore the underlying mechanisms linking math anxiety and math performance, with particular focus on the roles of working memory and math self-concept as potential mediators. The rationale for this work stems from the lack of validated instruments for measuring math anxiety in Swedish-speaking populations, despite the availability of several tools for English-speaking contexts. Additionally, there is limited knowledge about how math anxiety and its interaction with cognitive and self-belief factors influences math performance in young students—a critical period during which math anxiety begins to develop.

To address this gap, this thesis presents the validation of the Swedish Math Anxiety Rating Scale–Elementary (MARS-E), applied across three different Nordic samples. Working memory and self-concept in math were studied as mediators in the math anxiety–math performance relationship. A meta-analysis synthesised the general association between math anxiety and working memory into a mean effect size. It also examined the indirect effect of math anxiety on math performance via working memory. This mediating mechanism (working memory), as well as math self-concept, was later studied empirically in Swedish Grade 4–5 students.

The Swedish MARS-E, as well as the mediating processes in the math anxiety–math performance link, is discussed further in the subsequent chapters.

5.1 The Swedish MARS-E

Although substantial evidence supports the multidimensionality of the math anxiety construct (Henschel & Roick, 2017; Suinn et al., 1988; Wigfield & Meece, 1988), the hypothesised dimensionality varies across instruments. Consequently, the literature on math anxiety is fragmented, with numerous ways to measure the construct.

In this thesis, the Swedish MARS-E was developed and validated to provide a more contextually appropriate instrument for use with primary school students in Nordic settings. Study I supported a unidimensional model of math anxiety in Swedish-speaking samples. Study II confirmed this unidimensional structure in two culturally distinct Swedish-speaking samples. In contrast, the Finnish-speaking sample showed a two-factor structure, consistent with the original German MARS-E, suggesting that dimensionality may be context-dependent.

These inconsistencies, both in the literature and in Study II, suggest that the dimensionality of math anxiety may depend on contextual, developmental, or methodological factors. While a more complex conceptualisation of math anxiety might capture more nuanced effects, a unidimensional construct offers a straightforward and intuitive interpretation.

5.1.1 Validity Aspects of the MARS-E

Validity considerations are important across all studies in this thesis, but they are particularly crucial in Study I, where the Swedish MARS-E was examined. Given the existence of multiple validity frameworks (Newton & Baird, 2016), assessing a test's validity risks appearing arbitrary, and even open for cherry-picking frameworks. In a review by Hussey and Hughes (2020), 86% of 26 published scales were deemed satisfactory. However, when considering more comprehensive validity metrics, such as factor structure and invariance (among other metrics), only 4% demonstrated adequate validity. This finding highlights a widespread validity problem, potentially driven by the same factors underlying the replication crisis (Open Science Collaboration, 2015) and questionable research practices like p-hacking[10] (Masicampo & Lalande, 2012).

There are many potential threats to consider when assessing the validity of a test. Messick (1995) identified two major risks to the validation process: construct underrepresentation and construct irrelevant variance. These threats are particularly relevant here, given the context of the Swedish MARS-E, which was adapted from Henschel and Roick's (2017) two-dimensional scale. Results from Study I suggested that a unidimensional structure provided a better fit. This raises the possibility of construct underrepresentation, implying that the scale might be too narrowly focused, potentially omitting important aspects of the multidimensional construct.

Validity is often framed around three core aspects: *content, criterion* and *construct* validity. Construct validity is often the most demanding to assess. Authors like Cronbach and Messick have described validation as a continuous, never-ending process (Newton & Shaw, 2014). Validity is not a static property of a test but a continuous process of evaluation and refinement.

[10]p-hacking, misuse of data for finding significant results.

The studies in this thesis have primarily focused on construct validity, ensuring that the Swedish MARS-E captures math anxiety across different groups and timepoints. However, expanding the validation framework may be necessary if the instrument is further developed—for instance, by introducing a scoring guide. At present, there are no consequences regardless of individual's responses. If the instrument were ever to be further developed for new contexts, such as a screening tool, new claims would need to be specified to support its intended use, as proposed by the *Standards*. In such cases, *consequences* would become a key aspect to examine and would likely require additional work on predictive validity.

5.1.2 Validation Work in the Swedish MARS-E

Standards (AERA et al., 2014) identify five main aspects of validity: *test content, response processes, internal structure, relations to other variables,* and *consequences*. In this thesis, the validity framework outlined in the *Standards* is used to guide the validation of the Swedish MARS-E. The Swedish MARS-E underwent a rigorous validation process in Study I, which was further extended in Study II. Each aspect of validity and how they relate to the development of the Swedish MARS-E is described below:

Test content involves empirical and logical analyses to assess the extent to which a test's items are representative of the intended construct. For the Swedish MARS-E, this process began with the translation and adaptation of the German version. This initial step provided the foundation for both the test content as well as a reference point for the internal structure. The translation and adaptation process included expert reviews, and the subsequent test revisions were guided by iterative pilot testing. During this process, we examined both the results and feedback provided by teachers and students.

Response processes are more important for some tests than others. For example, a math test designed to test a certain type of reasoning would require both theoretical and empirical evidence showing how test-takers approach the items. This type of evidence, typically collected from individual responses, must be representative of all target groups. Notably, this aspect of validity partially overlaps with content validity. For the Swedish MARS-E, we asked selected children how they perceived the test and invited pilot samples—including both teachers and students—to provide feedback. The four-point Likert scale used in the Swedish MARS-E provides the test-taker with a straightforward response format, reducing the risk of misinterpretation. The discrepancy in dimensionality observed in Study II, may reflect different response styles across cultural groups.

Evidence based on *internal structure* examines whether the relationships among test items align with the intended construct, ensuring that items function as expected. This includes evaluating construct dimensionality, analysing item interrelationships, and assessing whether test components behave consistently across different groups. The Swedish MARS-E was evaluated using Confirmatory Factor Analysis to assess overall model fit, along with convergent and discriminant validity of the latent math anxiety factors. The analysis also investigated how the construct functioned across genders and timepoints. Studies I and II made substantial efforts to explore the internal structure of math anxiety, including invariance testing across gender, time, and cultural groups.

Relations to other variables, or criterion validity, assess how well test scores correlate with external measures, ensuring consistency with the intended construct. This includes evaluating convergent validity (strong correlations with related constructs) and discriminant validity (weak correlations with unrelated constructs). In Study I, multiple external measures were used to assess the criterion validity of the Swedish MARS-E. These variables were selected to reflect and distinguish the math anxiety construct for convergent and discriminant validity.

The *consequences* of test use extend beyond score interpretation and can impact individuals, institutions, and broader systems. While some consequences may align with the intended purposes of the tests, others may unintentionally affect areas such as fairness, educational practices, or public perception. It is important to evaluate these effects to ensure that the test use remains valid, ethical, and aligned with its intended goals.

When dealing with high-stakes tests, like the SATs or entrance exams, the consequence of testing is apparent. For the Swedish MARS-E, while direct consequences for individuals may not be evident, the broader implications for research are significant. This is particularly relevant in light of the replication crisis as highlighted by the Open Science Collaboration (2015), which documented major challenges in replicating previous research findings.

Given the wide range of instruments used to measure math anxiety–some multidimensional, some unidimensional, some longer or more context-specific–we cannot assume that they are interchangeable. This needs to be considered when interpreting results across studies, as instruments may differ both conceptually and psychometrically. Such differences have implications for both meta-analyses and theory development.

5.1.3 Cross-Cultural Comparisons

The cross-cultural comparability of the Swedish MARS-E was examined in Study II across three Nordic samples: Finnish-speaking students from Finland, Swedish-speaking students from Finland (Finland-Swedish), and Swedish-speaking students from Sweden. Using Confirmatory Factor Analysis, we found distinct factor structures depending on language context. Both Swedish-speaking samples showed better fit for a unidimensional model, while a two-factor structure was supported for the Finnish-speaking sample.

The items for the Swedish MARS-E were originally adapted from Henchel and Roick's (2017) instrument, which conceptualised math anxiety as a two-factor construct consisting of cognitive and affective components. The Swedish version included eight items from each dimension. With this in mind, it was the Swedish-speaking groups that diverged from the original conceptualisation. It should be noted that in the Swedish-speaking groups, a two-factor solution fit the data according to fit indices. However, the extremely high correlation between the factors raised concerns about discriminant validity. Given that a unidimensional construct of math anxiety is both theoretically and statistically justifiable, we opted for the more parsimonious solution, which also simplifies interpretation.

Measurement invariance analyses indicated that the Swedish MARS-E exhibited scalar invariance across gender in all three samples, suggesting the instrument functioned equivalently for boys and girls within each cultural context. Invariance was also supported across the two Swedish-speaking groups.

Correlation analyses revealed negative relationships between math anxiety and math performance across all three samples, generally stronger for arithmetic fluency than number processing. However, gender-specific patterns varied. Correlations were often more pronounced for boys than girls, especially for number processing. In the Swedish sample, gender differences in correlations were not significant ($p > 0.05$). In the Finland-Swedish sample, number processing correlated significantly stronger with math anxiety in the boys' group (Wald $\chi^2 = 3.54$, $p < 0.001$), whereas in the Finnish project, the affective factor showed a significantly stronger correlation with arithmetic fluency among boys (Wald $\chi^2 = 1.98$, $p = 0.048$). These findings highlight potential gender-specific vulnerabilities in how math anxiety affects performance, where specifically boys seem to experience larger adverse effects. However, these patterns are not consistent across cultural groups, even when measured with culturally invariant instruments.

Furthermore, gender-specific analyses demonstrated that girls reported higher overall math anxiety levels than boys across all three samples, where the largest gender difference originated from the Finland-Swedish sample (Cohen's d = 0.65). Interestingly, in the Finnish-speaking sample, no significant gender difference was found in the affective factor.

5.2 Mediators in the Anxiety–Performance Link

Various models of working memory have been proposed over the years (Shah & Miyake, 1999). This thesis adopted the Baddeley and Hitch framework as it is widely accepted with a strong empirical foundation (Baddeley, 2010), and aligns closely with the study design and instruments used in Studies III and IV. In Study III, we categorised measurements of working memory according to the Baddeley model and conducted subgroup analysis. Study IV used a latent variable approach, treating working memory tasks as indicators of a latent construct. This approach presents both strengths and considerations. On the one hand, the model fit the data well, with good fit indices and factor loadings of 0.5 and above, supporting the idea of a unitary yet multifaceted working memory system. On the other hand, it could be argued that the Baddeley model calls for a multi-factor structure, separating working memory's distinct components (phonological loop and visuospatial sketchpad).

Study III followed a more compartmentalised approach, aligning directly with the Baddeley and Hitch model, by treating the phonological and visuospatial components separately. In contrast, Study IV used a single latent factor, which aligns more with Engle and colleagues' (1999) unidimensional conceptualisation, describing working memory as:

"Working memory = Short-Term Memory + Central Executive" (p. 323).

Rather than being contradictory, these frameworks seem to complement each other. While Baddeley's model emphasises distinct functions of subsystems, Engle's model highlights individual differences in overall capacity and executive control. The latent variable approach in Study IV captures shared variance across tasks, providing a broad and comprehensive measure of working memory capacity. This approach also accounts for measurement error, thereby avoiding the limitations of assuming that single tasks provide error-free measurement.

In sum, Study IV grounds the instruments and tasks in the Baddeley and Hitch framework, while the statistical modelling also fits well with Engle et al.'s (1999) concept of using latent variables to capture individual differences. Notably, both conceptualisations seem to be rooted in empirical data derived from the similar tasks (e.g. operation span, various span tasks, and spatial tasks).

Study III and IV show clear links between math anxiety, math performance, and working memory. Working memory appears to, at least partially, mediate the math anxiety–math performance relationship. However, exactly how this mediation occurs is not clear from the included studies. Applying the Attentional Control Theory (Eysenck et al., 2007) to a math context offers a plausible explanation, proposing that math anxiety depletes attentional control—a resource required for effective math performance. Still, Studies III and IV have limitations in directly demonstrating these mechanisms. For instance, in Study IV, working memory and math anxiety were measured on the same day, but not simultaneously. As both constructs were treated as stable (trait), this assumption may not fully capture the participants' actual experiences. It is possible that individuals did not experience math anxiety during working memory testing. Nonetheless, correlations between the two were found in both Study IV and in the review in Study III.

This raises the possibility that Study IV may be affected by systematic confounders, such as general or test anxiety. Although working memory is often conceptualised as a trait variable with a stable maximum capacity, research has shown that this capacity can fluctuate due to sleep deprivation, stress, and other situational factors. This suggests that working memory also exhibits state-like properties (Engle, 2010). While this does not invalidate the trait conceptualisation, it complicates interpretation. Working memory capacity could appear reduced because it is partly occupied with intrusive, math-related thoughts (such as worry), or because of general state factors (e.g., acute stress) that temporarily impair capacity. Both processes produce similar outcomes, but the underlying mechanisms differ.

Self-concept in the anxiety – performance link

The data suggest that performance is negatively affected by anxiety, partly through cognitive interruptions (Eysenck et al., 2007). Individuals experiencing math anxiety may need to exert more effort to combat its impact on working memory, which they are more likely to do if they perceive the tasks as valuable. However, if their expectations success is low, they may disengage, even when the situation demands more effort rather than less (Wigfield & Eccles, 2000).

As established in an earlier chapter, self-concept is closely related to math achievement. Students with lower self-concept tend to show reduced achievement, both concurrently and over time (Marsh & Martin, 2011). Given that self-concept both affects and is affected by math anxiety

(Ahmed et al., 2012), it is reasonable to consider it as a mediator in the math anxiety–performance relationship. Study IV examined the roles of both working memory and self-concept in the math anxiety–math performance relationship. The main model in Study IV allowed for a direct comparison of working memory and math self-concept as mediators. The results suggested that both mediators were similarly important. This complements the directionality proposed in Lau et al.'s (2024) model and extends it by suggesting that self-concept may also mediate the reverse pathway, from anxiety to performance. This finding aligns with Expectancy-Value Theory, which posits ability beliefs as mediators between affective reactions and academic outcomes (Eccles & Wigfield, 2020).

5.3 Limitations

This thesis contributes to the development and validation of the Swedish MARS-E, as well as a deeper understanding of the mediators between math anxiety and math performance. Of course, no research is perfect—this thesis is no exception. With contributions come limitations, and several should be acknowledged.

First, the generalisability of the findings is somewhat limited. While Study II examined cross-cultural validity across three Nordic samples, these samples still share cultural and educational similarities, which may limit the applicability of the results to more diverse educational systems. Second, the content validity of the Swedish MARS-E, assessed in Study I, could benefit from qualitative research. Interviews might provide more nuanced insights into how students experience math anxiety and how item wording could be improved. This is particularly relevant since Study I did not meaningfully distinguish between the cognitive and affective factors in math anxiety. Qualitative methods could potentially clarify which mathematical contexts trigger state anxiety.

Third, mediation effects over time might be imprecise. Implementing autoregressive cross-lagged panel models could have better captured the temporal stability and directional influences among the variables. In addition, the constructs in Study IV were measured at a trait level, without controlling for participants' actual state anxiety at the time of testing. Addressing this would require more controlled testing environments and could benefit from physiological measures to assess state anxiety in real time.

Finally, a problem with measuring the effects of math anxiety on cognitive or performance variables lies in the variance overlap with related constructs. Study I demonstrated a high correlation between test anxiety and math anxiety, and it is reasonable to assume some overlap between general and math anxiety. However, when modelling these relationships, adding multiple anxiety-related variables alongside autoregressive paths into an already complex model, poses a risk of over-controlling for shared variance.

5.4 Looking Forward

The cognitive consequences of math anxiety are often based on the assumption that it disrupts working memory (Sidney et al., 2024). While Studies III and IV provide evidence that working memory partially mediates the anxiety–performance link, they fall short of capturing working memory in the moment when anxiety disrupts performance during mathematical problem solving. Naturally, there are ethical challenges associated with studies that attempt to measure in-the-moment disruptions, as inducing anxiety will cause mental discomfort. However, without carefully designed studies measuring in-the-moment (state) math anxiety, we are operating on assumptions about the student's state based on measures of the student's trait, leaving a gap in our understanding of how math anxiety impairs performance.

Regarding the validation process of tests, it is considered necessary to assess the social consequences (AERA et al., 2014). So far, the Swedish MARS-E has only been used in research settings for correlational studies, having no direct impact on the students, regardless of their score. However, as the Swedish MARS-E is already used in intervention studies, future need may include screening for math anxiety outside of research, given the reports of its impact on STEM courses and even career choices (Lau et al., 2024). Future work could explore the development of a scoring guide. Given the Swedish MARS-E's unidimensional structure, where both cognitive and affective items contribute to a single general factor, interpretation is relatively straightforward: higher scores reflect greater general math anxiety. Additional qualitative research could further clarify how students experience math anxiety. Such work could support the interpretation of scores and inform intervention-work.

5.5 Conclusion

This thesis deepens our understanding of math anxiety and its role in shaping students' mathematical development. Through psychometric, cross-cultural, meta-analytical, and longitudinal approaches, the work demonstrates that math anxiety is a measurable, meaningful construct, even in young learners, with real implications for working memory, self-beliefs, and academic outcomes. Beyond contributing to theoretical models, this research highlights the value of early identification and intervention. Addressing math anxiety is not only about reducing anxious thoughts and emotions, but also about supporting long-term academic development, and potentially influencing future educational and career paths.

References

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*.

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

Alloway, T. P., & Passolunghi, M. C. (2011). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences*, *21*(1), 133-137. https://doi.org/https://doi.org/10.1016/j.lindif.2010.09.013

Amsterdam, B. (1972). Mirror self-image reactions before age two. *Developmental psychobiology*, *5*(4), 297-305. https://doi.org/10.1002/dev.420050403

Asare, B., Arthur, Y. D., & Al-hassan, A.-M. (2025). Moderating effect of math anxiety and the mediating role of student self-efficacy on the nexus between cognitive awareness and student math performance. *Journal of Applied Research in Higher Education*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/JARHE-07-2024-0365

Ashcraft, M. H. (2002). Math anxiety: personal, educational, and cognitive consequences. *Curr. Dir. Psychol. Sci.*, *11*, 181–185. https://doi.org/10.1111/1467-8721.00196

Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychonomic bulletin & review*, *14*, 243-248. https://doi.org/ https://doi.org/10.3758/BF03194059

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

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829-839. https://doi.org/10.1038/nrn1201

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), R136-R140. https://doi.org/https://doi.org/10.1016/j.cub.2009.12.014

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological review*, *105*(1), 158-173.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47-89). Academic Press. https://doi.org/https://doi.org/10.1016/S0079-7421(08)60452-1

Balhinez, R., & Shaul, S. (2019). The Relationship Between Reading Fluency and Arithmetic Fact Fluency and Their Shared Cognitive

Skills: A Developmental Perspective [Original Research]. *Frontiers in psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01281

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and social psychology*, *51*(6), 1173-1182. https://doi.org/10.1037/0022-3514.51.6.1173

Baumeister, R. F. (2010). Self. In E. J. Finkel & R. F. Baumeister (Eds.), *Advanced social psychology: The state of the science* (pp. 139-176). Oxford University Press, Incorporated.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, *15*, 1-40.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (Second edition ed.). The Guilford Press.

Carey, E., Hill, F., Devine, A., & Szücs, D. (2016). The chicken or the egg? The direction of the relationship between mathematics anxiety and mathematics performance. *Front. Psychol.*, *6*, 1987. https://doi.org/10.3389/fpsyg.2015.01987

Carey, E., Hill, F., Devine, A., & Szűcs, D. (2017). The Modified Abbreviated Math Anxiety Scale: A Valid and Reliable Instrument for Use with Children [Original Research]. *Frontiers in psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00011

Chan, D. (2010). So why ask me? Are self-report data really that bad? In *Statistical and methodological myths and urban legends* (pp. 329-336). Routledge.

Cheung, M. W. (2014). Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. *Psychol Methods*, *19*(2), 211-229. https://doi.org/10.1037/a0032968

Ching, B. H.-H. (2017). Mathematics anxiety and working memory: Longitudinal associations with mathematical performance in Chinese children. *Contemporary Educational Psychology*, *51*, 99-113. https://doi.org/https://doi.org/10.1016/j.cedpsych.2017.06.006

Cipora, K., Artemenko, C., & Nuerk, H.-C. (2019). Different ways to measure math anxiety. In I. Mammarella, S. Caviola, & A. Dowker (Eds.), *Mathematics anxiety* (pp. 20-41). Routledge. https://doi.org/10.4324/9780429199981-2

Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: a practical guide to

Kane's framework. *Medical education.*, *49*(6), 560-575. https://doi.org/10.1111/medu.12678
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/https://doi.org/10.1016/j.jecp.2009.01.004
Decarli, G., Sella, F., Lanfranchi, S., Gerotto, G., Gerola, S., Cossu, G., & Zorzi, M. (2023). Severe Developmental Dyscalculia Is Characterized by Core Deficits in Both Symbolic and Nonsymbolic Number Sense. *Psychological Science*, *34*(1), 8-21. https://doi.org/10.1177/09567976221097947
Doz, E., Cuder, A., Pellizzoni, S., Granello, F., & Passolunghi, M. C. (2024). The interplay between ego-resiliency, math anxiety and working memory in math achievement. *Psychological Research*, *88*(8), 2401-2415. https://doi.org/10.1007/s00426-024-01995-0
Du, C., Qin, K., Wang, Y., & Xin, T. (2021). Mathematics interest, anxiety, self-efficacy and achievement: Examining reciprocal relations. *Learning and Individual Differences*, *91*, 102060. https://doi.org/https://doi.org/10.1016/j.lindif.2021.102060
Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 101859. https://doi.org/https://doi.org/10.1016/j.cedpsych.2020.101859
Eidlin Levy, H., & Rubinsten, O. (2021). Numbers (but not words) make math anxious individuals sweat: Physiological evidence. *Biological Psychology*, *165*, 108187. https://doi.org/https://doi.org/10.1016/j.biopsycho.2021.108187
Endler, N. S., & Kocovski, N. L. (2001). State and trait anxiety revisited. *Journal of Anxiety Disorders*, *15*(3), 231-245. https://doi.org/https://doi.org/10.1016/S0887-6185(01)00060-3
Engle, Randall W. (2010). Role of Working-Memory Capacity in Cognitive Control. *Current Anthropology*, *51*(S1), S17-S26. https://doi.org/10.1086/650572
Engle, R. W., Laughlin, J. E., Tuholski, S. W., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach [Article]. *Journal of Experimental Psychology: General*, *128*(3), 309-331. https://doi.org/10.1037/0096-3445.128.3.309

Eysenck, M. W. (1979). Anxiety, learning, and memory: A reconceptualization. *Journal of Research in Personality*, *13*(4), 363-385. https://doi.org/https://doi.org/10.1016/0092-6566(79)90001-1

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and Performance: The Processing Efficiency Theory. *Cognition and Emotion*, *6*(6), 409-434. https://doi.org/10.1080/02699939208409696

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion*, *7*, 336. https://doi.org/10.1037/1528-3542.7.2.336

Fang, J., Huang, X., Zhang, M., Huang, F., Li, Z., & Yuan, Q. (2018). The Big-Fish-Little-Pond Effect on Academic Self-Concept: A Meta-Analysis [Systematic Review]. *Frontiers in psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01569

Finell, J., Eklöf, H., Korhonen, J., & Jonsson, B. (2024). Reliability and validity evidence of the Swedish shortened mathematics anxiety rating scale elementary (MARS-E). *Discover Education*, *3*(1), 240. https://doi.org/10.1007/s44217-024-00348-8

Finell, J., Sammallahti, E., Korhonen, J., Eklöf, H., & Jonsson, B. (2022). Working Memory and its mediating role on the relationship of math anxiety and math performance: A meta-analysis. *Frontiers in psychology*, *12*, 798090. https://doi.org/10.3389/fpsyg.2021.798090

Fioriti, C. M., Pizzie, R. G., Evans, T. M., Green, A. E., & Lyons, I. M. (2025). Math anxiety and arithmetic learning: Evidence for impaired procedural learning and enhanced retrieval learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/https://dx.doi.org/10.1037/xlm0001453

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

Geary, D. C. (2013). Early Foundations for Mathematics Learning and Their Relations to Learning Disabilities. *Current Directions in Psychological Science*, *22*(1), 23-27. https://doi.org/10.1177/0963721412469398

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. https://doi.org/10.1080/00220973.1995.9943818

Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, *5*(10), 3-8. https://doi.org/10.3102/0013189x005010003

Gunderson, E. A., Park, D., Maloney, E. A., Beilock, S. L., & Levine, S. C. (2018). Reciprocal relations among motivational frameworks, math anxiety, and math achievement in early elementary school. *Journal of Cognition and Development*, *19*(1), 21-46. https://doi.org/https://doi.org/10.1080/15248372.2017.1421538

Guzmán, B., Rodríguez, C., & Ferreira, R. A. (2021). Longitudinal Performance in Basic Numerical Skills Mediates the Relationship Between Socio-Economic Status and Mathematics Anxiety: Evidence From Chile [Original Research]. *Frontiers in psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.611395

Hakkarainen, A., Väisänen, E., Aunio, P., Hellstrand, H., Laakso, M.-J., Laine, A., Räsänen, P., & Korhonen, J. (2025). Psychometric Evidence for the Functional Numeracy Assessment Dyscalculia Battery (FUNA-DB Screener). *European Journal of Psychological Assessment*, *0*(0), null. https://doi.org/10.1027/1015-5759/a000895

Hart, S. A., & Ganley, C. M. (2019). The Nature of Math Anxiety in Adults: Prevalence and Correlates. *J Numer Cogn*, *5*(2), 122-139. https://doi.org/10.5964/jnc.v5i2.195

Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! testing! one, two, three – Testing the theory in structural equation models! *Personality and Individual Differences*, *42*(5), 841-850. https://doi.org/https://doi.org/10.1016/j.paid.2006.10.001

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39-65. https://doi.org/https://doi.org/10.1002/jrsm.5

Hellstrand, H., Holopainen, S., Korhonen, J., Räsänen, P., Hakkarainen, A., Laakso, M.-J., Laine, A., & Aunio, P. (2024). Arithmetic fluency and number processing skills in identifying students with mathematical learning disabilities. *Research in Developmental Disabilities*, *151*, 104795. https://doi.org/https://doi.org/10.1016/j.ridd.2024.104795

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *J. Res. Math. Educ.*, *21*, 33–46. https://doi.org/10.2307/749455

Henschel, S., & Roick, T. (2017). Relationships of mathematics performance, control and value beliefs with cognitive and affective math anxiety. *Learning and Individual Differences*, *55*, 97-107. https://doi.org/10.1016/j.lindif.2017.03.009

Henschel, S., & Roick, T. (2020). The Multidimensional Structure of Math Anxiety Revisited. *European Journal of Psychological Assessment*, *36*(1), 123-135. https://doi.org/10.1027/1015-5759/a000477

Hitch, G. J., & Baddeley, A. D. (1976). Verbal Reasoning and Working Memory. *Quarterly Journal of Experimental Psychology*, *28*(4), 603-621. https://doi.org/10.1080/14640747608400587

Ho, H.-Z., Senturk, D., Lam, A. G., Zimmer, J. M., Hong, S., Okamoto, Y., Chiu, S.-Y., Nakazawa, Y., & Wang, C.-P. (2000). The affective and cognitive dimensions of math anxiety: A cross-national study. *Journal for research in Mathematics Education*, *31*(3), 362-379. https://doi.org/https://doi.org/10.2307/749811

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The abbreviated math anxiety scale (AMAS) construction, validity, and reliability. *Assessment*, *10*(2), 178-182. https://doi.org/https://doi.org/10.1177/1073191103010002008

Hussey, I., & Hughes, S. (2020). Hidden Invalidity Among 15 Commonly Used Measures in Social and Personality Psychology. *Advances in Methods and Practices in Psychological Science*, *3*(2), 166-184. https://doi.org/10.1177/2515245919882903

Justicia-Galiano, M. J., Martín-Puga, M. E., Linares, R., & Pelegrina, S. (2017). Math anxiety and math performance in children: The mediating roles of working memory and math self-concept. *British Journal of Educational Psychology*, *87*(4), 573-589. https://doi.org/10.1111/bjep.12165

Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, *50*(1), 1-73. https://doi.org/https://doi.org/10.1111/jedm.12000

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

Korhonen, J., & Räsänen, P. (2018). *Dimensionality of math anxiety and its longitudinal associations with math performance in adolescent students*. In 9th conference of the Nordic Research network on Special Needs Education in Mathematics (NORSMA) in Vaasa, Finland,

Kyttälä, M., & Björn, P. M. (2022). Mathematics Performance Profiles and Relation to Math Avoidance in Adolescence: The Role of Literacy Skills, General Cognitive Ability and Math Anxiety. *Scandinavian Journal of Educational Research*, *66*(7), 1221-1236. https://doi.org/10.1080/00313831.2021.1983645

Lau, N. T. T., Ansari, D., & Sokolowski, H. M. (2024). Unraveling the interplay between math anxiety and math achievement. *Trends in cognitive sciences*, *28*(10), 937-947. https://doi.org/10.1016/j.tics.2024.07.006

Lazarus, R. S. (2001). Relational meaning and discrete emotions. In *Appraisal processes in emotion: Theory, methods, research*. (pp. 37-67). Oxford University Press.

Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological reports*, *20*(3), 975-978.

Lindskog, M., Winman, A., & Poom, L. (2017). Individual differences in nonverbal number skills predict math anxiety. *Cognition*, *159*, 156-162. https://doi.org/https://doi.org/10.1016/j.cognition.2016.11.014

Liu, Y., Millsap, R. E., West, S. G., Tein, J. Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological methods*, *22*(3), 486. https://doi.org/10.1037/met0000075

Luttenberger, S., Sigrid, W., & and Paechter, M. (2018). Spotlight on math anxiety. *Psychology Research and Behavior Management*, *11*(null), 311-322. https://doi.org/10.2147/PRBM.S141421

Ma, X., & Xu, J. (2004). The causal ordering of mathematics anxiety and mathematics achievement: a longitudinal panel analysis. *Journal of adolescence*, *27*(2), 165-179. https://doi.org/https://doi.org/10.1016/j.adolescence.2003.11.003

Maldonado Moscoso, P. A., Castaldi, E., Arrighi, R., Primi, C., Caponi, C., Buonincontro, S., Bolognini, F., & Anobile, G. (2022). Mathematics and Numerosity but Not Visuo-Spatial Working Memory Correlate with Mathematical Anxiety in Adults. *Brain Sciences*, *12*(4), 422. https://www.mdpi.com/2076-3425/12/4/422

Maloney, E. A. (2016). Math anxiety: Causes, consequences, and remediation. In K. Wentzel & D. Miele (Eds.), *Handbook of motivation at school* (2nd ed., pp. 408-423). Routledge. https://doi.org/https://doi.org/10.4324/9781315773384

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. https://doi.org/10.1080/17470218.2010.533278

Maloney, E. A., & Beilock, S. L. (2012). Math anxiety: Who has it, why it develops, and how to guard against it. *Trends in cognitive sciences*, *16*(8), 404-406.

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. https://doi.org/10.1177/0956797615592630

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

Marsh, H. W. (1986). Verbal and Math Self-Concepts: An Internal/External Frame of Reference Model. *American educational research journal*, *23*(1), 129-149. https://doi.org/10.3102/00028312023001129

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on psychological science*, *1*(2), 133-163. https://doi.org/https://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, *81*(1), 59-77.

Marsh, H. W., Relich, J. D., & Smith, I. D. (1983). Self-concept: The construct validity of interpretations based upon the SDQ. *Journal of Personality and social psychology*, *45*(1), 173. https://doi.org/10.1037/0022-3514.45.1.173

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic Self-Concept, Interest, Grades, and Standardized Test Scores: Reciprocal Effects Models of Causal Ordering. *Child Development*, *76*(2), 397-416. https://doi.org/https://doi.org/10.1111/j.1467-8624.2005.00853.x

Masicampo, E. J., & Lalande, D. R. (2012). A peculiar prevalence of p values just below .05. *Quarterly Journal of Experimental Psychology*, *65*(11), 2271-2279. https://doi.org/10.1080/17470218.2012.711335

Messick, S. (1989). Validity. In L. Roberts (Ed.), *Educational Measurement* (3 ed., pp. 13-104). American Counsil on Education.

Messick, S. (1995). Validity of Psychological Assessment: Validation of Inferences From Persons' Responses and Performances as Scientific Inquiry Into Score Meaning. *The American psychologist*, *50*(9), 741-749. https://doi.org/10.1037/0003-066X.50.9.741

Mononen, R., Niemivirta, M., Korhonen, J., Lindskog, M., & Tapola, A. (2022). Developmental relations between mathematics anxiety, symbolic numerical magnitude processing and arithmetic skills from first to second grade. *Cognition and Emotion*, *36*(3), 452-472. https://doi.org/10.1080/02699931.2021.2015296

Mueller, C. E., & Winsor, D. L. (2016). Math and verbal academic self-concept: Subject-specificity across four distinctive groups of high ability adolescents. *Learning and Individual Differences*, *50*, 240-245. https://doi.org/https://doi.org/10.1016/j.lindif.2016.08.002

Namkung, J. M., Peng, P., & Lin, X. (2019). The relation between mathematics anxiety and mathematics performance among school-aged students: a meta-analysis. *Rev. Educ. Res.*, *89*, 459–496. https://doi.org/10.3102/0034654319843494

Newton, P. E., & Baird, J.-A. (2016). The great validity debate. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 173-177. https://doi.org/10.1080/0969594X.2016.1172871

Newton, P. E., & Shaw, S. D. (2014). *Validity in Educational and Psychological Assessment*. SAGE Publications Ltd.

Nyberg, L., & Eriksson, J. (2016). Working Memory: Maintenance, Updating, and the Realization of Intentions. *Cold Spring Harb Perspect Biol*, *8*(2), a021816. https://doi.org/10.1101/cshperspect.a021816

O'Leary, T. M., Hattie, J. A. C., & Griffin, P. (2017). Actual Interpretations and Use of Scores as Aspects of Validity. *Educational Measurement: Issues and Practice*, *36*(2), 16-23. https://doi.org/https://doi.org/10.1111/emip.12141

OECD. (2023). *PISA 2022 database*. Retrieved 20.12.2023 from https://www.oecd.org/en/data/datasets/pisa-2022-database.html#intro

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/doi:10.1126/science.aac4716

Orbach, L., Herzog, M., & Fritz, A. (2019). Relation of State- and Trait-Math Anxiety to Intelligence, Math Achievement and Learning Motivation. *Journal of Numerical Cognition*, *5*(3), 371-399. https://doi.org/10.5964/jnc.v5i3.204

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement Emotions and Academic Performance: Longitudinal Models of Reciprocal Effects. *Child Development*, *88*(5), 1653-1670. https://doi.org/https://doi.org/10.1111/cdev.12704

Pelegrina, S., Martín-Puga, M. E., Lechuga, M. T., Justicia-Galiano, M. J., & Linares, R. (2024). Role of executive functions in the relations of state- and trait-math anxiety with math performance. *Annals of the New York Academy of Sciences*, *1535*(1), 76-91. https://doi.org/https://doi.org/10.1111/nyas.15140

Pellizzoni, S., Cargnelutti, E., Cuder, A., & Passolunghi, M. C. (2022). The interplay between math anxiety and working memory on math performance: a longitudinal study. *Annals of the New York Academy of Sciences*, *1510*(1), 132-144. https://doi.org/https://doi.org/10.1111/nyas.14722

Plake, B. S., & Parker, C. S. (1982). The Development and Validation of a Revised Version of the Mathematics Anxiety Rating Scale. *Educational and Psychological Measurement*, *42*(2), 551-557. https://doi.org/10.1177/0013164482042002 18

Podlogar, N., Komidar, L., & Juriševič, M. (2024). Internal structure validity of the Self Description Questionnaire II and latent academic self-concept profiles of 9th grade students. *Learning and Individual Differences*, *116*, 102559. https://doi.org/https://doi.org/10.1016/j.lindif.2024.102559

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior research methods*, *40*(3), 879-891. https://doi.org/https://doi.org/10.3758/BRM.40.3.879

Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math anxiety: Past research, promising interventions, and a new interpretation framework. *Educational psychologist*, *53*(3), 145-164. https://doi.org/10.1080/00461520.2018.1447384

Räsänen, P., Aunio, P., Laine, A., Hakkarainen, A., Väisänen, E., Finell, J., Rajala, T., Laakso, M.-J., & Korhonen, J. (2021). Effects of Gender on Basic Numerical and Arithmetic Skills: Pilot Data From Third to Ninth Grade for a Large-Scale Online Dyscalculia Screener. *Frontiers in education*, *6*. https://doi.org/10.3389/feduc.2021.683672

Richardson, F. C., & Suinn, R. M. (1972). The mathematics anxiety rating scale: psychometric data. *Journal of counseling Psychology*, *19*(6), 551. https://doi.org/10.1037/h0033456

Sarason, I. G. (1957). Test anxiety, general anxiety, and intellectual performance. *Journal of Consulting Psychology*, *21*(6), 485.

Shah, P., & Miyake, A. (1999). Models of Working Memory: An Introduction. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 1-27). Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9781139174909.004

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-Concept: Validation of Construct Interpretations. *Review of educational research*, *46*(3), 407-441. https://doi.org/10.2307/1170010

Sidney, P. G., Scheibe, D. A., Zahrn, L., Brown, K. G. I., & Thompson, C. A. (2024). Developing Effective Interventions for Math Anxiety. *Current Directions in Psychological Science*, *0*(0), 09637214241300111. https://doi.org/10.1177/09637214241300111

Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex roles*, *50*, 241-252.
Sorvo, R., Kiuru, N., Koponen, T., Aro, T., Viholainen, H., Ahonen, T., & Aro, M. (2022). Longitudinal and situational associations between math anxiety and performance among early adolescents. *Annals of the New York Academy of Sciences*, *1514*(1), 174-186. https://doi.org/https://doi.org/10.1111/nyas.14788
Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., Tolvanen, A., & Aro, M. (2019). Development of math anxiety and its longitudinal relationships with arithmetic achievement among primary school children. *Learning and Individual Differences*, *69*, 173-181. https://doi.org/https://doi.org/10.1016/j.lindif.2018.12.005
St Omer, S. M., & Chen, S. (2023). Examining the dynamics of mathematics anxiety, perceived cost, and achievement: A control-value theory approach. *Contemporary Educational Psychology*, *73*, 102169. https://doi.org/https://doi.org/10.1016/j.cedpsych.2023.102169
Suinn, R. M., Taylor, S., & Edwards, R. W. (1988). Suinn mathematics anxiety rating scale for elementary school students (MARS-E): Psychometric and normative data. *Educational and Psychological Measurement*, *48*(4), 979-986.
Swanson, L., & Kim, K. (2007). Working memory, short-term memory, and naming speed as predictors of children's mathematical performance. *Intelligence*, *35*(2), 151-168. https://doi.org/https://doi.org/10.1016/j.intell.2006.07.001
Träff, U., Östergren, R., Skagerlund, K., & Skagenholt, M. (2025). Mental arithmetic skill development in primary school: The importance of number processing abilities and general cognitive abilities. *Journal of Experimental Child Psychology*, *252*, 106155. https://doi.org/https://doi.org/10.1016/j.jecp.2024.106155
Vukovic, R. K., Kieffer, M. J., Bailey, S. P., & Harari, R. R. (2013a). Mathematics anxiety in young children: Concurrent and longitudinal associations with mathematical performance. *Contemporary Educational Psychology*, *38*(1), 1-10. https://doi.org/https://doi.org/10.1016/j.cedpsych.2012.09.001
Vukovic, R. K., Roberts, S. O., & Green Wright, L. (2013b). From Parental Involvement to Children's Mathematical Performance: The Role of Mathematics Anxiety. *Early Education and Development*, *24*(4), 446-467. https://doi.org/10.1080/10409289.2012.693430
Wang, L., & Yu, Z. (2023). Gender-moderated effects of academic self-concept on achievement, motivation, performance, and self-

efficacy: A systematic review [Systematic Review]. *Frontiers in psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1136141

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.

Wang, Z., Rimfeld, K., Shakeshaft, N., Schofield, K., & Malanchini, M. (2020). The longitudinal role of mathematics anxiety in mathematics development: Issues of gender differences and domain-specificity. *Journal of adolescence*, *80*, 220-232. https://doi.org/https://doi.org/10.1016/j.adolescence.2020.03.003

Wigfield, A., & Eccles, J. S. (2000). Expectancy–Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, *25*(1), 68-81. https://doi.org/https://doi.org/10.1006/ceps.1999.1015

Wigfield, A., & Meece, J. L. (1988). Math anxiety in elementary and secondary school students. *Journal of educational psychology*, *80*(2), 210.

Wiklund-Hörnqvist, C., Jonsson, B., Korhonen, J., Eklöf, H., & Nyroos, M. (2016). Untangling the Contribution of the Subcomponents of Working Memory to Mathematical Proficiency as Measured by the National Tests: A Study among Swedish Third Graders [Original Research]. *Frontiers in psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01062

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

Zhang, J., Chiu, M. M., & Lei, H. (2023). Achievement, self-concept and anxiety in mathematics and English: A three-wave cross-lagged panel study. *British Journal of Educational Psychology*, *93*(1), 56-72. https://doi.org/https://doi.org/10.1111/bjep.12539

Živković, M., Pellizzoni, S., Mammarella, I. C., & Passolunghi, M. C. (2023). The relationship betweens math anxiety and arithmetic reasoning: The mediating role of working memory and self-competence. *Current Psychology*, *42*(17), 14506-14516. https://doi.org/10.1007/s12144-022-02765-0