DOCTORAL THESIS

Navigating Methodological Challenges

A Quantitative Exploration of Students' Mathematics Self-Beliefs

Erik Bergqvist

Mathematics Education

Luleå University of Technology

Department of Health, Education and Technology
Division of Education and Languages

Doctoral thesis in Mathematics Education

Navigating Methodological Challenges: A Quantitative Exploration of Students' Mathematics Self-Beliefs

Erik Bergqvist

September 2024

Erik Bergqvist

Navigating Methodological Challenges:

A Quantitative Exploration of Students' Mathematics Self-Beliefs

Doctoral thesis in Mathematics Education, September 2024

Supervisors: Maria Johansson and Timo Tossavainen

Luleå University of Technology

Department of Health, Education and Technology

Division of Education and Languages

97187 Luleå

Abstract

Mathematics self-beliefs influence students' self-image, classroom behavior, engagement, and motivation. However, current research faces challenges due to overlapping constructs, which can lead to high correlations and ambiguous findings. For instance, while mathematics self-efficacy is often defined broadly in educational research, task-specific mathematics self-efficacy refers specifically to confidence in completing particular tasks. To enhance the credibility of predictions regarding mathematics achievement, it is crucial to distinguish between different types of general mathematics self-beliefs. This thesis aims to provide new insights into these beliefs by refining methodologies for their measurement.

An exploratory factor analysis of upper-secondary student data, utilizing both quantitative and qualitative reasoning, revealed a clear distinction between mathematics self-concept and generalized mathematics self-efficacy beliefs. A key difference between these two constructs is that the latter significantly influences student engagement. Additionally, the identified factor structure included two constructs related to mathematics anxiety. One noteworthy finding is that the worries of providing incorrect answers in the classroom significantly contribute to gender differences and diminish students' agentic engagement. Overall, these findings underscore the importance of removing variables with substantial intercorrelations from the analysis, provided this procedure is supported by qualitative reasoning. The findings of this thesis support a pragmatic approach to investigating mathematics self-beliefs, ultimately offering new insights into the significant role of autonomy-supportive teaching in shaping students' general mathematics self-beliefs.

Keywords: Mathematics Self-efficacy, Mathematics Self-concept, Operationalization challenges, Exploratory Factor Analysis, Student Engagement and Motivation

Acknowledgement

I wish to begin by expressing my gratitude to my supervisors for their invaluable guidance and unwavering support during my doctoral studies. I would also like to extend my thanks to my colleagues at the Division of Education and Languages, especially to my PhD student colleagues at the doctoral school, PROFS. The discussions, whether during formal seminars or over coffee breaks, kept me engaged and inspired. A special thank you goes to my colleague and friend, C-A. Our extensive and thought-provoking discussions on a range of topics have significantly broadened my perspective.

I would like to express my heartfelt gratitude to my wife and our four children for their support during this academic journey. Their presence has been a constant source of strength, particularly during challenging times that have impacted our family deeply. I want to give a special thanks to my wife, my closest friend, whose encouragement motivated me to pursue research after nearly twenty years of teaching mathematics and physics at the upper-secondary school level. Your unwavering support means the world to me, and I am deeply grateful to have you by my side. Finally, I want to express my thanks to my parents and mother-in-law for their support, which has allowed me to devote extra time to refining this thesis.

Erik
Luleå, September 2024

Preface

The inspiration for my research stems from my interest in programming and its potential to enhance student self-beliefs in mathematics. At its core, programming is similar to mathematical problem-solving, as both require precise thinking and insights that extend beyond mere technical skills. However, my research focus shifted significantly.

Firstly, I observed a lack of clarity in defining mathematics self-efficacy beliefs, as well as diverse interpretations of how programming activities are integrated into mathematics education. These discrepancies could present challenges in both analyzing data and implementing programming tasks in educational settings, which drew my attention and interest for further research. Additionally, my experience as a mathematics teacher provided me with insight into the significance of students' engagement with the subject. Consequently, my research shifted focus to exploring how to credibly measure mathematics self-efficacy beliefs despite methodological concerns and the relationship between these beliefs and mathematics teaching.

Before beginning my doctoral studies, my proficiency in statistics was relatively limited, particularly concerning the topics explored in this thesis. Although I have a solid foundation in mathematics and experience as a mathematics teacher, I would like to emphasize that I am not a statistician. Consequently, this research required significant effort to apply statistical methods appropriately. However, this situation also presented some advantages, as it allowed me to leverage my prior teaching experience to address critical issues and approach the research from a pragmatic perspective.

List of Papers

The following papers are referred to in the text by the corresponding Roman numerals (I-IV).

Paper I Bergqvist, E., Tossavainen, T., & Johansson, M. (2020). An analysis of high and low intercorrelations between mathematics self-efficacy, anxiety, and achievement variables: A prerequisite for a reliable factor analysis. *Education Research International, 2020*. https://doi.org/10.1155/2020/8878607

Paper II Bergqvist, E. (2024). Relations Between Mathematics Self-Efficacy and Anxiety Beliefs: When Multicollinearity Matters. *The Journal of Experimental Education*, 1-21. https://doi.org/10.1080/00220973.2024.2338545

Paper III Bergqvist, E. (2022). An inquiry of different interpretations of programming in conjunction with mathematics teaching. In G. A. Nortvedt, N. F. Buchholtz, J. Fauskanger, M. Hähkiöniemi, B. E. Jessen, M. Naalsund, H. K. Nilsen, G. Pálsdóttir, P. Portaankorva-Koivisto, J. Radii, J. Ö. Sigurjónsson, O. Viirman, & A. Wernberg (Eds.), *Bringing nordic mathematics education into the future. Proceedings of norma 20 the ninth nordic conference on mathematics education*. Svensk förening för MatematikDidaktisk Forskning - SMDF.

Paper IV Bergqvist, E. (2024). Predicting Student Engagement: The Role of Generalized Mathematics Self-Efficacy [Submitted]

The articles included in the thesis are reprinted with permission from the publishers and the co-authors.

List of abbreviations

The following table describes various abbreviations and acronyms used throughout this thesis.

Abbreviation	Meaning	Page(s)
CFA	Confirmatory Factor Analysis	37, 52, 54, 58, 59
CT	Computational Thinking	34, 49
CTT	Classical Test Theory	53, 55, 71
EFA	Exploratory Factor Analysis	18, 39, 48, 51, 52, 53, 57, 58, 59, 61, 72, 73
IRT	Item Response Theory	52, 53, 54, 71
MCMC	Markov Chain Monte Carlo	55, 56, 64, 66
MSEAQ	Mathematical Self-Efficacy and Anxiety Questionnaire	37, 39, 59, 97, 103, 105
ROPE	Region of Practical Equivalence	65, 66
SDT	Self-Determination Theory	10, 11, 22, 23, 77

Contents

Part I

Framework and Findings Synthesis

Introduction 1

During my time as a mathematics teacher, I frequently observed that students who struggled with mathematics often exhibited lower levels of confidence and persistence. Those who doubted their mathematical abilities commonly displayed behaviors consistent with these negative beliefs. Students who perceived themselves as poor at mathematics often reinforced this perception by avoiding study before tests, focusing on other activities, or seeking reassurance from peers with similar self-views. Furthermore, when I provided feedback that challenged these perceptions, students sometimes resisted and offered counterarguments, attempting to maintain their existing self-concepts. This phenomenon, where self-perception is difficult to change, has been recognized in research since the 1970s (Markus, 1977). For instance, when students succeeded on a test, they frequently dismissed my positive feedback, attributing their success to luck or claiming that the test was too easy. This situation not only perpetuated their negative beliefs about mathematics but also hindered my ability to support them effectively, ultimately affecting their engagement in class.

The example above, drawn from my teaching experience, illustrates the significant role that a teacher's actions can play in influencing students' motivation to learn mathematics (Reeve, 2018). This finding aligns with previous research emphasizing the connection between teachers' instructional methods, student engagement, and motivation (Appleton et al., 2008; Connell & Wellborn, 1991; H. Jang et al., 2016; Reeve & Shin, 2020; Ryan & Deci, 2020; Skinner & Belmont, 1993; Skinner et al., 2008). However, empirically investigating students' self-beliefs in mathematics presents several methodological challenges. This thesis addresses many of these challenges and offers suggestions for enhancing the credibility of quantitative research in educational settings. Additionally, it explores the significance of these beliefs within the context of mathematics education.

1.1 Rationale and background

The interplay between students' self-beliefs in mathematics and contextual factors, such as teachers' instructional methods, significantly influences their engagement and motivation, encompassing aspects such as focus, curiosity, and problem-solving skills. For instance, motivated students are more likely to prefer deep learning strategies over superficial approaches (Reeve, 2012; Skinner et al., 2009). Furthermore, motivated students actively engage in their learning by asking questions that aid their understanding and facilitate progress—a concept known as agentic engagement (Reeve, 2012). This proactive involvement prompts adjustments in teachers' instructional methods, thereby fostering greater support for student autonomy (H.-R. Jang et al., 2024; Reeve et al., 2004). Teachers who demonstrate care and adaptability can enhance students' enthusiasm for learning and nurture their sense of autonomy in the learning process (Reeve & Shin, 2020). In essence, the dynamic interaction between teachers and students highlights the crucial role of instructional practices in shaping students' individual self-beliefs. Moreover, this underscores the importance of considering the mathematics teaching context when examining the relationship between students' self-beliefs in mathematics and their engagement with the subject.

Previous research on students' motivation, as evidenced by student engagement (Skinner et al., 2009), highlights significant variations in the instruments used for measurement. Particularly noteworthy is the assessment of general mathematics self-beliefs, including generalized mathematics self-efficacy, mathematics self-concept, and mathematics anxiety. A lack of clarity regarding the distinctions between these concepts leads to unclear operationalization and complicates the interpretation of research findings. Consequently, diverse instruments may unintentionally assess the same construct (Bong & Skaalvik, 2003; Lee et al., 2020; Marsh et al., 2019). While this thesis primarily emphasizes quantitative research, it is important to acknowledge that operationalization issues are equally critical in qualitative research, despite the differences in approaches across research paradigms. Fundamentally, operationalization involves the same process: adapting the overall research objective for investigation (L. Cohen et al., 2018).

Specifically, in quantitative research, unclear operationalization can lead to extreme multicollinearity in data, which arises from exceptionally high correlations among the observed variables. This creates challenges in accurately interpreting certain statistical analyses, including factor analysis and regression analysis (Field et al., 2013; Tabachnick & Fidell, 2013), as demonstrated by Marsh et al. (2004). Moreover, some research has demonstrated that when comparing results between countries, the constructs of mathematics self-concept and mathematics self-efficacy used in the Programme for International Student Assessment (PISA) 2003 and 2012 cycles might be unreliable (e.g., Ding et al., 2023). This suggests that interpreting results from cross-country comparisons of these constructs should be approached with caution. Another challenge in assessing mathematics self-beliefs involves the philosophy of measurement. First and foremost, accurately measuring these beliefs is complex due to their indirect and subjective nature. Consequently, they may lack a quantitative structure and may not be measurable using metric scales (Michell, 2012).

However, a widely accepted method for addressing measurement challenges involves defining measurement as a procedure that entails the "assignment of numerals to objects or events according to a rule" (Stevens, 1946, p. 677). This definition can be further refined to include the process of selecting appropriate measurement scales, such as nominal, ordinal, or metric. However, this approach has potential pitfalls that are frequently overlooked. The issue primarily arises from the fact that metric scales are generally more desirable in data analysis compared to nominal or ordinal scales, as they allow the use of conventional statistical methods. For example, although Likert scales are inherently ordinal rather than metric, Pearson's correlation is often misapplied in social and behavioral research when estimating correlations between variables measured with Likert scales (Chen & Popovich, 2002). The practical significance of this issue, however, remains debatable (see Norman, 2010).

In addressing the aforementioned challenges, it is crucial to consider the difficulties in conceptually separating generalized mathematics self-efficacy from mathematics self-concept (Bong & Skaalvik, 2003). Consequently, methodologies that effectively address the complexities of defining and measuring these concepts are required. Such methodologies can enhance the depth of analysis and improve the interpretation of results. For example, in educational research, mathematics self-beliefs are often used to predict students' mathematics performance. However, it may be necessary to reevaluate these predictions. Furthermore, the significant correlations identified in multiple studies (e.g., Marsh et al., 2019, 2004; Pajares & Miller, 1994) suggest that even well-established scales for measuring general mathematics self-efficacy beliefs require a thorough examination of the observed variables. Such an analysis is essential to avoid potential errors caused by multicollinearity among highly correlated or overlapping variables. Specifically, extreme multicollinearity among observed variables in factor analysis can complicate the interpretation of the factor structure (Field et al., 2013; Tabachnick & Fidell, 2013) and limit the consistency of results in future studies (Rockwell, 1975). Notably, despite this, educational researchers often overlook this issue, possibly due to the misconception that extreme multicollinearity is problematic only in small-scale studies.

In summary, prior research has highlighted the challenges in defining and measuring general mathematics self-beliefs. Consequently, the relationship between these self-beliefs and student engagement remains underexplored. This underscores the need for exploratory studies, even on previously validated scales, to generate new insights and guide future research in this domain.

1.2 Focus of research

The conclusions drawn in this thesis are based on four original publications listed on page (ix). The research seeks to enhance the understanding of students' general mathematics self-beliefs. Its objective is to improve the methodology used to investigate these beliefs. This objective is pursued by evaluating two specific aspects of the research:

- *Operationalization challenges* of general mathematics self-beliefs and the proper application of factor analysis to mitigate extreme multicollinearity in data.
- *Significance of teaching context* in analyzing students' general mathematics self-beliefs, and its implications for understanding the relationship between general mathematics self-beliefs and student engagement.

The findings of this thesis are based on empirical data gathered from students, including their self-evaluations of general mathematics self-beliefs within the context of mathematics instruction.

1.3 Research questions

The specific research questions vary among the studies (see Papers I-IV), with each study aiming to explore one or more of the following main research questions.

1. *What quantitative methodology is suitable for analyzing general mathematics self-beliefs when high multicollinearity is present in survey data?* This question is significant because there is a prevalent lack of clarity in operationalizing mathematics self-beliefs, often leading to exceedingly high multicollinearity in the collected data.
2. *What latent factors manifest as a result of applying this methodological approach?* This question aims to investigate how employing factor analysis, coupled with pre-treatment for extreme multicollinearity, influences the factor structure in a sample of students.
3. *What is the relationship between these factors and student engagement in mathematics teaching?* This question aims to investigate the impact

of employing factor analysis, combined with pre-treatment for extreme multicollinearity, on understanding the interplay between students' general mathematics self-beliefs, and their engagement.

Papers I and II are connected to the exploration of the first and second research questions, respectively, while Paper IV addresses the third research question. Paper III clarifies the research context in which the third research question is investigated. It examines diverse perspectives on programming in mathematics teaching, highlighting the challenges involved in integrating programming into mathematics education. Given the exploratory nature of the empirical studies, their primary goal is to uncover new insights and issues for further investigation, with a particular focus on students' general self-beliefs in mathematics.

1.4 Structure of the thesis

This thesis is organized into two primary sections: a research framework with a synthesis of findings, and four individual publications. The research framework integrates these publications into a unified thesis and situates them within a broader research context. The second chapter introduces the conceptual framework employed. Chapters three and four detail the research design and methodological choices made for the empirical studies. The final two chapters synthesize and discuss the findings from the individual publications, addressing their relevance, limitations, and implications for mathematics education.

Conceptual framework 2

In this chapter, I will begin by exploring the fundamental concepts central to this thesis. I will then discuss how these concepts interrelate and are influenced by external factors, such as the teaching context. Finally, I will present a theoretical framework that integrates these concepts within the overarching theme of engagement. This model aims to illustrate the development of student motivation and its significance within the classroom setting.

2.1 Motivation

In the introduction, I briefly outlined the importance of motivation in mathematics education and its relationship with both personal beliefs and external factors. In the following sections, I will delve into the concept of motivation and the factors that influence it.

A fundamental description of motivation is that it serves as the driving force behind our actions toward achieving goals (Baumeister, 2016; Ryan & Deci, 2020). Motivation is influenced by external factors (context) and three primary motives: our needs, cognition, and emotions (see Figure 2.1). However, motivation is an internal experience that cannot be directly measured. Therefore, evaluating motivation relies on observable indicators such as behavior and self-reports (Reeve, 2018). For instance, in this thesis, motivation is evaluated by analyzing self-reported data on student engagement and general mathematics self-beliefs.

2.1.1 The driving forces of motivation

Our needs are essential for life and well-being, encompassing not only physiological needs but also three basic psychological needs: autonomy, competence, and relatedness (Ryan & Deci, 2020). Autonomy involves a sense of ownership over one's actions, which is enhanced by experiences of value, interest, and joy. The need for competence relates to the feeling of mastery or the sense of being capable of succeeding and making progress. Relatedness involves a sense of belonging and connection with others (Ryan & Deci, 2020). Accordingly, motivated students perceive themselves as competent individuals with agency, believing in multiple paths to success and growth, along with a profound sense of belonging and connection with their peers. This sense of relatedness flourishes in a classroom characterized by adaptability, mutual respect, and care, where optimal challenges and positive feedback are provided without reliance on external controls such as punishments or rewards (Reeve, 2012; Ryan & Deci, 2020).

Another influence on motivation is *implicit motives*, which arise from social experiences (see Figure 2.1). These implicit motives include the need for achievement, affiliation (connection with others), and power. They profoundly influence our thoughts, emotions, behaviors, and life choices (Reeve, 2012). For instance, individuals with a strong need for achievement find fulfillment in overcoming challenges, while those driven by affiliation or power seek close relationships or opportunities for societal impact. Additionally, our cognitive processes, including our plans and beliefs, significantly contribute to our motivational framework. Our emotions, manifested through feelings, physical responses, and actions, also play a pivotal role in driving our motivation. Lastly, the effects of the social context and external events on motive status are significant.

Understanding motivation in its entirety is complex. Consequently, numerous theories exist, each focusing on different aspects such as needs, cognition, or emotions. These theories complement one another and offer varied perspectives on motivation (see Figure 2.1). For instance, some theories examine decision-making and responses to failure (see Situated Expectance-Value Theory in J. S. Eccles & Wigfield, 2020), while others emphasize goal-setting (see e.g., Dweck, 1986). Among the most comprehensive frameworks for understanding motivation is Self-Determination Theory (SDT), which elucidates the key factors influencing intrinsic motivation (Ryan & Deci, 2020).

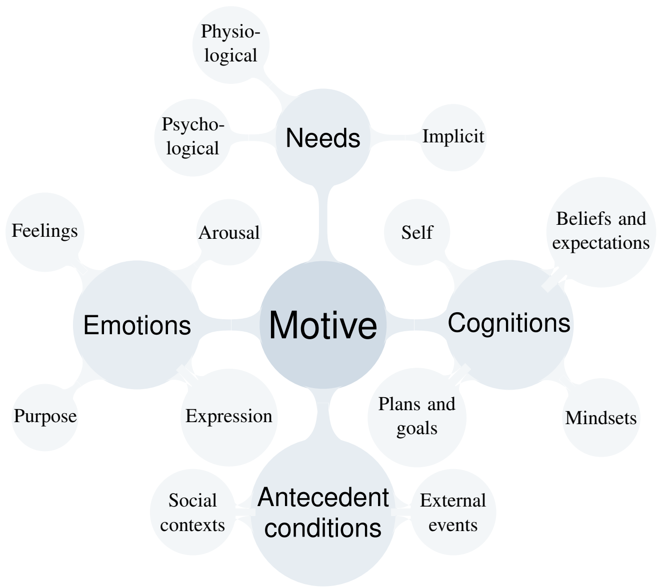

Figure 2.1. Framework for motivation (Adapted from Reeve, 2018, p. 13)

2.1.2 Types of motivation

Within SDT, motivation is classified into two primary types: extrinsic and intrinsic. When individuals lack the drive to engage in activities, this state is referred to as amotivation. Intrinsic motivation involves pursuing tasks due to genuine enjoyment derived from them. On the other hand, extrinsic motivation is a multifaceted construct influenced by external factors such as rewards or pressure. The degree of autonomy, or the extent to which individuals feel in control of their actions, can vary. Consequently, extrinsic motivation is further categorized into external, introjected, identified, and integrated regulation (Ryan & Deci, 2020).

External regulation represents the least autonomous form of motivation, where individuals engage in an activity solely to attain a reward or evade punishment. In contrast, *introjected regulation* often manifests as ego-involvement, where individuals comply with external demands by seeking approval from others, engaging in social comparisons, or striving to uphold a specific self-image or status (Ryan & Deci, 2020).

Moving to a more autonomous form, *identified regulation* occurs when individuals consciously acknowledge the value of the external rule and internalize it, making the regulation feel like a personal choice and commitment. Finally, *integrated regulation* represents the most autonomous form of extrinsic motivation, wherein an individual recognizes and identifies with the value of the activity, finding it consistent with one's own core interests and values (Ryan & Deci, 2020).

2.2 Competence beliefs

Belief in one's abilities is essential for sustaining motivation and plays a significant role in our understanding of motivation (Pajares, 1996) and the quality of engagement (Linnenbrink & Pintrich, 2003; Zimmerman, 2000). Motivational theories often use terms such as self-efficacy and self-concept to describe competence. Additionally, belief in one's abilities is linked to a fundamental psychological need: the sense of competence. For example, when individuals perceive themselves as proficient in a task, they are typically more interested and motivated (Reeve, 2012; Ryan & Deci, 2020). However, as outlined in the introduction, precisely delineating these beliefs and establishing reliable measurement methodologies can be particularly challenging, especially in the context of general mathematics self-beliefs.

Furthermore, beyond self-perceived abilities, competence can also be interpreted in terms of subject-specific skills. For example, it may refer to proficiency in technology, often termed digital competence, or to abilities in mathematics, referred to as mathematical competence. These distinct perspectives further underscore the importance of competence beliefs in motivating individuals and highlight their significance, particularly in the context of mathematics teaching.

2.2.1 Various facets

The perception of competence as varying among individuals is a crucial perspective due to its significant impact on motivation. For instance, when competence is perceived as a skill, improvements typically lead to positive outcomes, as higher competence is usually advantageous. However, if competence is viewed subjectively, a high level of competence may not always be beneficial. Individuals who are perceived as highly competent can sometimes evoke jealousy in others, which may lead to negative emotions (Weiner, 2005).

Our understanding of competence can be framed within the context of attribution theory (see Weiner, 2005), which categorizes competence beliefs based on our views about intellectual abilities (Dweck & Leggett, 1988). If abilities are perceived as innate and unchangeable, competence is seen as fixed and linked to inherent traits—a concept known as a fixed mindset. Conversely, if competence is viewed as a quality that results from effort and perseverance, it is seen as something that can be developed and enhanced over time, referred to as a growth mindset. Weiner (2005) illustrates why beliefs about competence are particularly crucial in the context of mathematics education. For instance, consider the difference between perceiving a highly skilled car mechanic and an individual with high mathematical competence. Car mechanics are often seen as competent when they adapt to new car technologies; if they do not, their competence may be perceived as diminished. In contrast, mathematical competence is generally regarded as static and attributed to inherent traits, with fewer situations in which this competence is perceived as fluctuating.

In summary, understanding mathematical competence from a personal perspective is crucial not only for comprehending why students might lack motivation in mathematics but also for its potential influence on teaching approaches. Hypothetically, if competence were perceived solely as equivalent to ability, teaching would likely focus primarily on content and the enhancement of students' mathematical skills. However, if competence is also seen as a personal attribute, teaching would be more naturally framed within a supportive environment where students learn to view mistakes as an integral part of the learning process. Furthermore, the concept of *self* is important because it can be understood both as an agent of action (e.g., "I will do my homework") and as a reflection of our beliefs about abilities and traits (Bandura, 2008). The role of the self in shaping beliefs about competence is particularly apparent when examining concepts such as self-efficacy and self-concept. To clarify why distinguishing between these beliefs can be challenging, it is crucial to first define and explore the concepts of self-concept and self-efficacy.

2.2.2 Mathematics self-concept

On a general conceptual level, self-concept is both hierarchical and multidimensional. This implies that self-concept encompasses various domain-specific facets, such as those related to mathematics and other non-academic areas. Furthermore, self-concept is hierarchical, meaning that these domain-specific facets operate at different levels of abstraction. General inferences about the self are situated at the top, while more specific inferences pertain to academic (e.g., mathematics) and non-academic subdomains (Shavelson et al., 1976). Mathematics self-concept, in particular, refers to an individual's confidence in learning mathematics, performing well in mathematics classes, and achieving high scores on mathematics tests (Reyes, 1984). It is influenced by personal perceptions as well as the perceptions of others, thereby contributing to the overall self-image. This includes how individuals perceive themselves in mathematics relative to other subjects (Marsh & Shavelson, 1985; Shavelson et al., 1976).

Since the initial proposal of the self-concept model (Shavelson et al., 1976), various refinements have been introduced based on empirical research (see Arens et al., 2021). For instance, it has been suggested that the hierarchical structure of the self-concept becomes more complex with age (Marsh & Shavelson, 1985). Additionally, verbal and mathematics self-concepts have been found to be almost uncorrelated (Marsh, 1986). Despite this lack of correlation, individuals with strong mathematical skills often exhibit high verbal skills as well (Marsh et al., 1988). To provide a theoretical rationale for these observations, Marsh and colleagues (Marsh, 1986; Marsh et al., 1988) proposed an internal/external reference model, which suggests that self-concept is shaped by both external and internal comparisons. This led to the development of a new model for the academic self-concept structure, known as the Marsh/Shavelson model (Marsh, 1990b). The internal/external reference model has received support from several research studies. For example, Marsh and Hau (2004) utilized PISA data to demonstrate its applicability across 26 countries.

External comparison involves comparing oneself with peers, such as other students in the same classroom, while internal comparison assesses performance across different subjects. Both processes contribute to forming a person's self-concept, which is both a cause and an effect of achievement, as described in the reciprocal effects model (Marsh, 1990a; Marsh & Yeung, 1997; Wu et al., 2021). The Big-Fish-Little-Pond Effect, a phenomenon substantiated by numerous studies (e.g., Marsh, 1987; Marsh & Parker, 1984), refers to the tendency of students to exhibit a lower self-concept when comparing themselves with perceived more talented peers and a higher self-concept when comparing themselves with less competent peers.

Integrating both internal and external comparisons is essential for understanding the very low correlation between verbal and mathematics self-concept. For instance, as illustrated by Marsh and Shavelson (1985), a student who believes they are not proficient in either mathematics or English compared to their peers (external comparison) may, upon self-reflection, recognize that they are better at mathematics than at English (internal comparison). Consequently, comparing themselves to others may result in a positive correlation between their verbal and mathematics self-concept. On the other hand, internal comparisons might lead to the belief that strength in one area implies weakness in the other, resulting in a negative correlation between verbal and mathematics self-concept. Thus, research indicates that when both internal and external comparisons are accounted for, the relationship between verbal and mathematics self-concept tends to be weak.

In summary, social comparison and evaluations of prior experiences significantly impact students' mathematics self-concept, which, in turn, affects their behavior and reinforces their self-perception (Bong & Skaalvik, 2003; Shavelson et al., 1976; E. M. Skaalvik & Skaalvik, 2002).

2.2.3 Mathematics self-efficacy

Self-efficacy is a key concept introduced by Albert Bandura within the framework of social cognitive theory (Bandura, 1977). This theory plays a crucial role in understanding human behavior, emphasizing the interaction between personal, behavioral, and social factors, and highlighting human agency as a central element (Schunk & Pajares, 2005; Usher & Pajares, 2008). Consequently, it underscores the importance of individuals' active involvement in their own development. According to Bandura (1982), "Perceived self-efficacy is concerned with judgments of how well one can execute courses of action required to deal with prospective situations" (p. 122). Furthermore, self-efficacy predicts behavioral changes and is distinct from outcome expectations, which represent the confidence in one's ability to act successfully to produce a certain outcome (Bandura, 1977). Self-efficacy beliefs are future-oriented perceptions regarding one's ability to perform specific tasks, with mathematics self-efficacy specifically referring to confidence in completing mathematical tasks (Pajares, 1996). Efficacy expectations are shaped by past experiences, including prior achievements, vicarious experiences from observing models, verbal persuasion, and emotional states (Bandura, 1982). These self-efficacy beliefs significantly influence behavior, choices, effort, resilience, emotional responses, and cognitive processes (Pajares, 1996; Schunk & Pajares, 2005).

Instead of assessing mathematics self-efficacy beliefs at the task-specific level as proposed by Albert Bandura, researchers sometimes use a more general level of specificity. Marsh et al. (2019) refers to this broader perception as generalized mathematics self-efficacy, which lacks specific criteria for successful performance. Consequently, individuals must apply a frame of reference to evaluate this self-belief. For instance, a task-specific mathematics self-efficacy assessment might be: "I am confident that I can successfully complete the following equation", whereas a generalized mathematics self-efficacy assessment might be: "I believe I can perform well on mathematics tests". Nevertheless, the empirical similarities between generalized mathematics self-efficacy and mathematics self-concept (Marsh et al., 2019) pose challenges for operationalization and have implications for data analysis (see Papers I and II).

In conclusion, individuals shape their mathematics self-efficacy beliefs by observing others perform tasks, particularly when they perceive similarities with those individuals. Additionally, verbal persuasion and emotional reactions—such as stress, anger, fear, and joy—also influence mathematics self-efficacy, as these emotions are interpreted as indicators of one's capability to succeed in a task. For instance, negative thoughts about one's abilities and feelings of nervousness can lower efficacy expectations, potentially creating a cycle of negative thinking that may intensify mathematics anxiety.

2.2.4 Mathematics anxiety

Previous research supports a strong correlation between mathematics self-concept and mathematics anxiety (e.g., Goetz et al., 2010; Pajares & Miller, 1994). This correlation might be attributed to the close link between emotions and motivation, as emotions serve as a type of motive (Reeve, 2018). Conceptually, emotions can be defined as biological reactions that guide actions and convey non-verbal signals to others (Reeve, 2018). Specifically, anxiety encompasses behavioral, cognitive, and affective components (Zeidner & Matthews, 2005). Although anxiety conceptually resembles fear, it differs in that fear typically involves a clearly defined threat, whereas anxiety is characterized by a more general sense of worry or tension, often related to future uncertainties (Reeve, 2018).

Mathematics anxiety is characterized by bodily responses associated with feelings of worry and helplessness, which hinder performance in mathematical problem-solving situations (Ashcraft, 2002). Both test anxiety and mathematics anxiety are conceptually linked to evaluation anxiety, which negatively impacts mathematical performance by disrupting cognitive processes in working memory (Ashcraft, 2002). Anxiety can be understood in terms of trait-related or state-related characteristics, with features typically being temporal rather than situational. For instance, test anxiety may be situational, arising before a specific exam, or it may be a recurring emotion related to testing scenarios (Pekrun, 2006; Zeidner & Matthews, 2005)

2.2.5 Operationalization challenges

As briefly outlined in the introduction, operationalizing generalized mathematics self-efficacy beliefs is challenging because distinguishing them from mathematics self-concept beliefs is difficult (Bong & Skaalvik, 2003; Marsh et al., 2019). This ambiguity may contribute to conflicting evidence regarding gender differences (Alves et al., 2016; Devine et al., 2012; Morán-Soto & González-Peña, 2022; Pajares & Graham, 1999; F. Xie et al., 2019). Additionally, this lack of clarity can lead to significant issues related to multicollinearity in the data, which is particularly problematic in Exploratory Factor Analysis (EFA) and regression analyses. Significant multicollinearity within factor analysis complicates the interpretation of factors, and addressing this issue can enhance the consistency of future research findings (Field et al., 2013; Rockwell, 1975; Tabachnick & Fidell, 2013).

Numerous guides are available for conducting EFA (e.g., Beavers et al., 2013; Costello & Osborne, 2005; Field et al., 2013; Tabachnick & Fidell, 2013; Yong & Pearce, 2013). A common challenge in this process is extreme multicollinearity, which arises when observed variables exhibit very high intercorrelation, indicating redundancy or overlapping information (Tabachnick & Fidell, 2013). While a certain degree of intercorrelation between variables is essential for EFA, excessive intercorrelation can hinder the ability to distinguish each variable's unique contributions, thus compromising the interpretability of the resulting factors (Field et al., 2013).

Addressing this issue is particularly important because the predictive power of self-efficacy beliefs on outcomes such as academic achievement depends on two critical factors: the specificity with which these beliefs are defined and the temporal alignment between the measurement of these beliefs and the occurrence of the outcomes (Bong & Skaalvik, 2003; Pajares, 1996). Consequently, conflating factors may diminish the interpretability of the predictive power of self-efficacy beliefs. For instance, task-specific mathematics self-efficacy has been shown to predict test-specific scores more accurately than overall school grades (Bandura, 2005; Marsh et al., 2005; Pajares, 1996). This issue is particularly relevant in the field of motivation research, where the use of numerous terms may lead to overlapping concepts. Specifically, the operationalization of generalized mathematics self-efficacy, mathematics self-concept, and mathematics anxiety can sometimes be conflated, leading to what are known as jingle-jangle fallacies (Bong & Skaalvik, 2003; Lee et al., 2020; Marsh et al., 2019). Jingle-jangle fallacies occur when the same term is used to represent different constructs (the jingle fallacy) or when different terms are used to represent the same construct (the jangle fallacy). Consequently, some scholars have suggested merging certain constructs to address this concern (see Pekrun, 2024).

However, in certain cases, it may be more appropriate to differentiate between closely related constructs rather than merging them (Pekrun, 2024), as the definitions of these concepts can vary significantly. For example, an individual may exhibit strong self-confidence in understanding the material of a specific mathematics course, indicating high generalized mathematics self-efficacy, while simultaneously perceiving themselves as less skilled in mathematics overall, reflecting a low mathematics self-concept. This phenomenon can be attributed to the fact that mathematics self-concept is conceptually more closely related to social comparison and past experiences than mathematics self-efficacy beliefs (Bong & Skaalvik, 2003).

In conclusion, mathematics self-efficacy should be interpreted cautiously to avoid confusion with generalized mathematics self-efficacy (e.g., “I am convinced that I can even understand the most difficult content in math”) and mathematics self-concept (e.g., “In math, I am a talented student”). However, Marsh et al. (2019) has expressed concern about whether mathematics self-concept and generalized mathematics self-efficacy can be empirically differentiated from one another. Similar concerns have also been raised regarding the differentiation between mathematics self-concept and anxiety (Klee et al., 2022), as well as expectations of success and academic self-concept (J. S. Eccles & Wigfield, 1995, 2002). Comparable issues exist among several terms used within motivational research (see Pekrun, 2024).

2.3 Student engagement

In educational contexts, *engagement* typically denotes students’ active participation in learning activities, signifying a high level of involvement. Although motivation and engagement are closely related, they represent distinct concepts. However, student engagement behavior is often seen as a visible indicator of motivation, making motivation more readily observable during classroom instruction (Connell & Wellborn, 1991; Ryan & Deci, 2020; Skinner & Belmont, 1993).

Researchers generally agree that student engagement is multifaceted (Fredricks et al., 2004), although its conceptualization may vary depending on the context (Appleton et al., 2008). Nevertheless, within educational research, it is beneficial to conceptualize student engagement as comprising behavioral, emotional, and cognitive dimensions, as well as their opposite, disengagement (Connell & Wellborn, 1991; Fredricks et al., 2004; Skinner et al., 2009). While these dimensions are often considered separable, it is crucial to acknowledge their interconnectedness (Fredricks et al., 2004; H. Jang et al., 2016; Skinner et al., 2009).

The interconnection between student engagement and personal self-beliefs is well-documented. Previous research highlights the mediating role of these perceptions in shaping the predictive influence of contextual elements, such as perceived autonomy support from teachers (e.g., Wang et al., 2017). Additionally, mathematics self-concept and mathematics self-efficacy are considered crucial in determining the quality of student engagement (Linnenbrink & Pintrich, 2003; Zimmerman, 2000). For instance, research consistently indicates a strong correlation between self-efficacy beliefs and student engagement (Linnenbrink & Pintrich, 2003; D. P. Martin & Rimm-Kaufman, 2015; Patrick et al., 2007; Sakiz et al., 2012).

2.3.1 Behavioral, emotional and cognitive engagement

Teachers most easily observe students' behavioral and emotional engagement. Behavioral engagement includes demonstrating effort and persistence, such as maintaining concentration and attention. Emotional engagement is evident when students show interest, joy, and enthusiasm. Conversely, disaffection is marked by both behavioral and emotional disengagement. This includes signs of passivity, withdrawal from activities, and expressions of frustration, anxiety, and boredom (Skinner et al., 2008, 2009).

Fredricks et al. (2004) defines cognitive engagement as involving both "thoughtfulness and the willingness to exert the effort necessary to comprehend complex ideas and master difficult skills" (p. 60). The relationship between self-regulated learning (e.g., "It was easy for me to establish goals for learning the material that will be on this exam") and cognitive engagement (e.g., "I put together ideas or concepts and drew conclusions that were not directly stated in course materials") is evident when examining the items in the scales used to measure these constructs (see Greene, 2015). This relationship is further supported by the definitions outlined in this thesis. Specifically, a cognitively engaged student employs learning strategies that facilitate deep learning (e.g., Connell & Wellborn, 1991; Reeve, 2018), while a cognitively disengaged student typically relies on less organized learning strategies (e.g., Elliot et al., 1999; H. Jang et al., 2016).

2.3.2 Agentic engagement

The commonly accepted definition of student engagement encompasses three key components: behavioral, emotional, and cognitive. While these aspects are broadly acknowledged as important, Reeve (2012) has highlighted a limitation: these dimensions largely emphasize how students respond to teacher-directed instructional activities, neglecting the proactive and transformative roles that students themselves play. To address this issue, Reeve suggested adding a fourth dimension to the concept of student engagement—the agentic dimension. This addition to student engagement has been explored in subsequent studies by Reeve (2013) and H. Jang et al. (2016).

Further, Reeve (2013) propose that "agentic engagement can be viewed not just as a student's contributions into the flow of instruction but also as an ongoing series of dialectical transactions between student and teacher" (p. 580). This form of engagement manifests in the classroom when a student actively participates by asking questions, providing feedback to enhance the learning environment for themselves and their peers, and expressing their needs and desires.

2.3.3 Linking the teaching context

The teaching context, including teachers' instructional practices, plays a crucial role in fostering student motivation (Skinner & Belmont, 1993; Skinner et al., 2008, 2009). Furthermore, recent research has supported the transformative role of student engagement on teachers' behavior, showing a connection between students' agentic engagement and teachers' instructional practices (H.-R. Jang et al., 2024). This finding can be attributed to the definition of agentic engagement, which, according to Reeve (2013), is both transactional and dialectical.

In the transactional aspect, students' engagement affects and shapes the actions of teachers delivering instruction, while teachers' actions, in turn, influence and shape students' engagement. In the dialectical aspect, when students proactively ask questions and engage in communication, it prompts adjustments in teachers' instructional methods, thereby increasing support for student autonomy (Reeve et al., 2004). Additionally, the teaching approach employed by educators significantly affects student-initiated engagement, creating a reciprocal feedback loop that fosters continuous improvement. This is supported by SDT, which suggests that an individual's motivational style significantly influences the motivation, emotions, learning, and performance of others (Ryan & Deci, 2020). The term *teaching motivating style* refers to the core attitude of teachers as they engage students in learning activities (Reeve & Shin, 2020). This style ranges from a responsive approach—autonomy-supportive, where teachers adjust to students' viewpoints, needs, and preferences—to a more controlling approach, where teachers dictate students' thoughts, emotions, and actions. Essentially, the relationship between agentic engagement and teachers' support for autonomy is mutually reinforcing (H.-R. Jang et al., 2024; Reeve et al., 2022).

Despite the significant role of teachers' motivational styles in influencing students' motivation, other contextual factors, such as the overall school culture and conditions, also play a crucial role (Pelletier & Rocchi, 2016). This dependency is explained by the fact that when teachers face limitations on their own freedom necessary for fostering students' learning and development—essentially, a lack of autonomy, competence, and relatedness satisfaction—they are less capable of addressing their students' basic psychological needs (Deci & Ryan, 2016). In the context of teaching, enhancing perceived autonomy support can be achieved by delivering mathematics instruction that is both engaging and meaningful for students (Tsai et al., 2008). Furthermore, previous research has established a connection between cognitive engagement and the teaching context. Specifically, students' cognitive engagement increases when they participate in problem-solving tasks and activities that encourage algorithmic thinking (Liu et al., 2023; Xing et al., 2023). This topic is discussed in Paper III regarding programming. This finding is particularly pertinent to Study II (see Chapter 4), where participants work on problems that are both mathematically complex and contextually realistic.

2.4 An integrative framework of motivation

A framework that brings together the concepts used in this thesis is the self-system model of motivational development. This model illustrates the interactions among student engagement, personal beliefs, and the teaching context, and it has proven especially valuable in understanding motivational processes in the classroom (Connell & Wellborn, 1991; Skinner & Belmont, 1993; Skinner et al., 2008, 2009).

Several reasons support the adoption of this model. Firstly, as previously discussed, this thesis adopts a perspective grounded in SDT, emphasizing the importance of individuals' psychological needs in understanding motivation and engagement (Ryan & Deci, 2020). This perspective defines an autonomously motivated student as one whose basic psychological needs are satisfied. In general, SDT has significant implications for education, as it helps teachers understand and foster not only individuals' intrinsic motivation for learning but also the resulting engagement (Reeve, 2012). Secondly, this model effectively illustrates the positive effects of engagement on both students and teachers. By demonstrating how increased student engagement can lead to greater satisfaction for teachers, the model highlights the reciprocal benefits of fostering engagement in the classroom. Additionally, since motivation is inherently subjective, personal beliefs about competence and contextual influences act as facilitators of engagement rather than merely as indicators (Skinner, 2016). Lastly, evaluating student engagement is often more practical than assessing motivation because engagement can be mea-

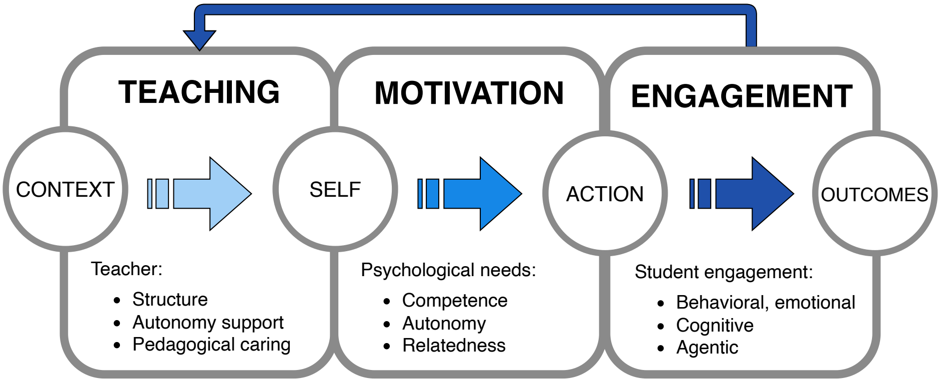

Figure 2.2. The self-system motivational development model within teaching context (Adapted from Skinner et al., 2008, p. 768)

sured through observable behaviors such as attentiveness, participation, and problem-solving.

The model (see Figure 2.2) focuses on four higher-order constructs: *Context*, *Self*, *Action*, and *Outcomes*. The term *Context* encompasses interactions with peers, parents, and teachers, the learning tasks and activities carried out in the classroom, and the methods of teaching employed. The model illustrated in Figure 2.2 highlights the importance of supportive teacher-student interactions. The term *Self* includes constructs such as competence beliefs, sense of belonging, beliefs about the value, relevance, and utility of schooling (see J. S. Eccles & Wigfield, 2020), and goal orientation (see e.g., Dweck, 1986). The term *Action* refers to the active participation, or lack thereof, exhibited by students in the classroom. It includes observable behaviors that indicate the degree of motivation students display and serves as an indicator of whether their basic psychological needs are being met within the learning environment. Finally, the term *Outcomes* refers to the results of engagement, such as learning and achievement.

In accordance with this model, when the teaching context addresses the student's psychological needs for autonomy, competence, and relatedness, it enhances student engagement. This model proves to be a valuable resource for understanding how various elements, including both contextual and personal factors such as beliefs about competence, can positively or negatively impact engagement and potentially lead to disengagement (Skinner et al., 2008). However, the process depicted in Figure 2.2 is not strictly linear; rather, the influences on motivational development are interconnected and reciprocal (H.-R. Jang et al., 2024; H. Jang et al., 2016). As a result, motivation creates an iterative loop that sustains the process. For instance, a caring, adaptable, and supportive teaching style can further motivate students to learn (H.-R. Jang et al., 2024; Reeve & Shin, 2020). This supportive environment, in turn, leads to enhanced knowledge and skills, potentially fostering interest and encouraging the use of deep learning strategies. These strategies may prompt more student-initiated questions, which, if the teacher responds receptively, can promote an even more supportive teaching style. Consequently, this cycle influences changes in student engagement (H.-R. Jang et al., 2024; Reeve et al., 2022).

The model illustrated in Figure 2.2 represents an adaptation of the original framework proposed by Skinner (2008). This revised model introduces agentic engagement as a new, fourth dimension of student engagement. Additionally, an arrow in the figure indicates the impact of engagement on teaching and underscores the reciprocal effects of student engagement. In summary, the model posits that when students achieve success in completing tasks or developing skills, it enhances their sense of competence and mathematics self-beliefs. This satisfaction of their basic psychological needs subsequently leads to higher levels of engagement (Connell & Wellborn, 1991; Furrer & Skinner, 2003; Reeve, 2013, 2012; Ryan & Deci, 2020; Skinner et al., 2008).

2.5 Significance of gender

While this thesis does not explicitly focus on gender disparities, acknowledging their relevance is crucial due to their profound influence on mathematics education. Stereotypes, such as the belief that boys outperform girls in mathematics, can negatively influence behaviors and personal beliefs within the mathematics classroom. This can, in turn, contribute to the underrepresentation of women in STEM (Science, Technology, Engineering, and Mathematics) fields. Therefore, investigating the validity of these stereotypes remains essential (Hyde, 2014).

2.5.1 Competence beliefs

Gender differences in mathematics self-efficacy (Pajares & Miller, 1994), mathematics self-concept (Guo et al., 2015; S. Skaalvik & Skaalvik, 2004), mathematics anxiety (Devine et al., 2012; Morán-Soto & González-Peña, 2022; Pajares & Kranzler, 1995; F. Xie et al., 2019), and general mathematics self-efficacy (Zander et al., 2020) have been observed in various studies. However, conflicting findings exist, particularly regarding mathematics anxiety and generalized mathematics self-efficacy (Alves et al., 2016; Pajares & Graham, 1999). For example, research by Tossavainen et al. (2021) reveals conflicting findings regarding first-year engineering students across three Nordic universities. The study identified no statistically significant gender differences in general mathematics self-efficacy. However, based on the conceptual framework outlined in this chapter, the term *mathematics self-concept* more precisely captures the essence of the construct, compared to *general mathematics self-efficacy*, which is less specific. Furthermore, this distinction is important because the applied measure includes items typically found in scales designed to assess mathematics self-concept.

Hyde (2005) posits that males and females exhibit substantial similarities on most psychological variables, which is known as the gender similarity hypothesis. For instance, although lower mathematics self-beliefs among females compared to males are often reported, social context likely influences these results (Else-Quest et al., 2010; Hyde, 2005). Moreover, the influence of gender differences may attenuate or disappear when controlling for additional variables, such as prior achievement (Pajares, 2005). Consequently, the presence of a gender disparity appears to depend on the inclusion of specific variables within regression models, highlighting the importance of addressing operationalization issues, as they can significantly impact the predictability of outcomes.

Despite evidence showing that males and females generally achieve similar levels of mathematical performance (Else-Quest et al., 2010; Hyde et al., 1990), a significant gender gap persists in terms of mathematics self-concept. Watt (2005) attributes this disparity to the observation that females' mathematics self-concept more accurately reflects their actual performance, while males may have a less realistic self-concept, leading them to engage in mathematical activities beyond their true capabilities. Additionally, research indicates that females are more likely to attribute their perceived lack of mathematical ability to failure, while males are more inclined to credit their success to their innate mathematical talent (J. Eccles et al., 1984). Goetz et al. (2013) further posits that this discrepancy might be linked to the stereotype that males are better at mathematics than females.

2.5.2 Student engagement

As previously discussed, gender differences play a significant role in the study of competence beliefs in mathematics. They are equally important for understanding student engagement. Previous research suggests that female students generally exhibit higher levels of both behavioral and emotional engagement compared to their male counterparts (Furrer & Skinner, 2003; Hofer et al., 2022; Lietaert et al., 2015; Skinner et al., 2008). Moreover, the disparity in behavioral engagement between genders may be partly attributed to the level of autonomy support provided by teachers (Lietaert et al., 2015). In contrast, research on gender differences in cognitive engagement has yielded inconsistent results. Some studies report no significant gender differences in cognitive engagement when prior personal goal orientation is controlled for (Young, 1997), while others indicate that females demonstrate higher levels of cognitive engagement, particularly in the use of strategies (Patrick et al., 1999; Zimmerman & Martinez-Pons, 1990).

Methodological framework 3

The aim of this chapter is to discuss the philosophical approach to measurement and research design.

3.1 Philosophical context

As outlined in the introduction, quantitative research presents numerous challenges, one of which is the philosophical approach to measurement. Despite its critical importance, the role of measurement is frequently underestimated in research (Kline, 2020). This issue is deeply intertwined with core questions of ontology and epistemology.

In the philosophical discourse on research methodologies, it is essential to acknowledge the inherent tension between quantitative and qualitative researchers. This tension reflects a broader conflict between pure positivism and pure interpretive constructivism (Johnson & Onwuegbuzie, 2004). Each paradigm advocates for the research methods it deems most effective, which can lead to a certain degree of dogmatism or reluctance to integrate methods from rival paradigms (Johnson & Onwuegbuzie, 2004). While some degree of dogmatism among researchers can be beneficial, as it prevents the premature dismissal of theories (Popper, 1963), this dogmatism becomes problematic when researchers focus exclusively on supporting existing theories and neglect potential refutations. As a result, incorrect ideas may lead to failure, which overlooks the importance of practical application and fails to fully recognize the role of success in scientific practice (Putnam, 1974). Moreover, dogmatism is significantly influenced by our self-perception and the ways in which we interpret observations and conceptualize the world. These processes are shaped by our underlying ontological and epistemological perspectives. This highlights the necessity for researchers to adopt a critical stance, which enables them to remain receptive to revising or even rejecting their theories when warranted, while still maintaining a reasonable level of confidence in their current theoretical frameworks (Lakatos, 1973; Popper, 1963).

As discussed in the previous chapter, measuring psychological attributes, such as mathematics self-beliefs, presents significant challenges because these attributes cannot be assessed directly like physical characteristics. Additionally, the indirect measurement of these constructs is complicated by potential issues with self-awareness (Reeve, 2018). Both traditional positivism and constructivism reject the notion that mental phenomena, such as beliefs, have the same ontological status as physical phenomena (Maxwell & Mittapalli, 2010). Positivists, in particular, argue that mental constructs are merely tools for predicting outcomes and do not have a direct connection to reality. This perspective may be influenced by a bias toward the physical sciences in understanding reality. Consequently, mental states are frequently considered less real because they are dependent on the mind (Maul, 2013).

Maul (2013) suggests that traditional perspectives on reality in quantitative psychology research, which are borrowed from the physical sciences, may have developed as a way to distance the field from pseudoscience. Nevertheless, Popper (1963) contends that research on psychological constructs bears some resemblance to pseudoscience, as theories in psychology often fail to explain behaviors that fall outside their theoretical boundaries. Despite this, these theories are meaningful and scientific, provided that researchers are willing to reject them when confronted with contradicting evidence (Popper, 1963). Furthermore, distinguishing between scientific and non-scientific methodologies is more relevant than differentiating between science and pseudoscience (Lakatos, 1973).

In educational research, particularly when examining constructs such as self-beliefs in mathematics, the methods for obtaining evidence differ significantly from those used for physical attributes. Unlike the natural sciences, where methods from physics are often directly applicable, social sciences face unique challenges. Specifically, isolating and studying a social phenomenon in isolation is frequently impractical. Consequently, knowledge generated in the social sciences can actively influence the system being studied. For instance, when a theoretical framework within this system evolves, individuals may adapt to this new understanding, thereby altering their behavior. This reciprocal interaction between theory and the studied system, as mentioned above, is more apparent in social science research compared to studies in the natural sciences (A. Chalmers, 2013). Furthermore, the interaction between the measurement tool and the object being measured can affect the object itself (Mari et al., 2023b).

In conclusion, the ontological perspective of empirical realism, which asserts that objects exist independently of the observer, seems less relevant in the context of educational research. Alternative interpretations of realism, as outlined by Maul (2013), offer more nuanced perspectives that avoid the rigid assumptions of empirical realism. One of these alternative views will be discussed in the following section.

3.2 The pragmatic perspective

Within the contrasting paradigms of positivism and constructivism, researchers often adhere exclusively to either quantitative or qualitative data. This tendency, as previously discussed, can present significant challenges. The existing tension between these paradigms may lead researchers to uncritically adopt the ontological and epistemological assumptions embedded within their chosen approach. However, by integrating elements from both perspectives, researchers can develop a pragmatic methodology that more effectively addresses the challenges related to objectivity and epistemology inherent in educational research.

The pursuit of conducting objective and value-neutral research is often endorsed. However, the feasibility of achieving this ideal depends on whether the focus is on ontology or epistemology. In general, research outcomes are influenced by subjective judgments (Peshkin, 1988; Polanyi, 1958). Researchers and their studies are inevitably shaped by their conceptual frameworks, with concepts, beliefs, and theories influencing their perception and interpretation of the world (Mari et al., 2023b; Polanyi, 1958). While it is important for researchers to maintain confidence in their theories, it is equally crucial to critically evaluate the assumptions underlying their chosen methodologies and to remain open to revising their beliefs as necessary (Hammersley & Traianou, 2012). In addressing concerns related to empirical realism, as discussed in the previous section and emphasized by scholars such as Mari et al. (2023b) and Guyon et al. (2018), this thesis adopts a pragmatic approach to measurement. This approach aligns with the model-based realist perspective on measurement, as outlined by Mislevy (2018). While a full exploration of this framework is beyond the scope of this thesis, it is important to note that it views reality socio-cognitively through multiple perspectives. These perspectives, including the conceptual framework, methodology, and research context, shape our results. From a pragmatic standpoint on measurement, psychological attributes are considered to possess both objective and intersubjective properties, reflecting a philosophical stance that integrates elements of pragmatism and realism. This perspective aligns with the form of realism advocated by philosopher Hilary Putnam (Mari et al., 2023b).

Specifically, psychological attributes are understood as manifestations of complex interactions among internal cognitive processes within the brain, as conceptualized within specific contexts (Guyon et al., 2018). Consequently, the construct being measured is assumed to exist within the particular social context in which the measurement occurs. Adopting a pragmatic-realism approach should not be confused with adopting a postmodern perspective. While mental attributes are understood within the context of our social practices, they are considered real entities (Guyon et al., 2018). This implies that, although these attributes are subjectively experienced—indicating ontological subjectivity—they can still be understood objectively from an epistemic standpoint. Thus, knowledge about an attribute can be considered objective if the method of acquiring this knowledge is independent of any particular opinion, regardless of the observer's identity (Maul, 2013). Furthermore, this pragmatic approach addresses the challenge of operationalization by emphasizing an interpretation of the results within the specific research context. For example, extensive multicollinearity may arise when respondents interpret multiple questionnaire items in similar or overlapping ways. The severity of this issue can vary based on the respondents and their interaction with the research context.

In conclusion, this thesis defines psychological attributes subjectively, from a pragmatic perspective. Nevertheless, these attributes remain measurable, as measurement itself is recognized as a social practice (Guyon et al., 2018; Mari et al., 2023b). This approach underscores the importance of the research context, which serves as a cornerstone for interpreting the results of this thesis. For example, as discussed later in this chapter, this perspective significantly influences my research, particularly regarding aspects such as generalization, the interpretation of statistical findings, and research design.

3.3 The research design

The research design employed in the empirical studies of this thesis resembles cross-sectional research by providing snapshots of students' mathematics self-beliefs. However, it diverges from traditional cross-sectional approaches due to its pragmatic orientation. Specifically, it integrates elements characteristic of case study methodology and the planning inherent in action research design (L. Cohen et al., 2018).

A case study can employ various research methodologies, such as action research, surveys, and naturalistic inquiry (L. Cohen et al., 2018). Typically, it utilizes an empirical approach to explore complex and ambiguous situations where the boundaries between the phenomenon being studied and its context are not clearly defined (Yin, 2017). Consequently, experimental research, which aims to isolate phenomena as much as possible, does not align with this definition (Yin, 2017). In the field of educational research, case studies often involve investigating specific phenomena from the researcher's established ontological and epistemological perspectives. These studies frequently address complex scenarios intimately connected to the teaching environment, thus fitting the definition of a case study (Yin, 2017). Furthermore, my experience as a teacher provides both an outsider's and an insider's perspective on the research. Therefore, the action research planning framework outlined by L. Cohen et al. (2018) is useful for clarifying my research process. Similar to case study methodology, action research incorporates various research strategies. However, it typically involves carefully monitored interventions designed to enhance practice, with their effects assessed through a combination of action and reflection.

During my time as a mathematics teacher, I reflected on how my role as an educator influenced students' motivation and its relationship to their self-beliefs in mathematics. Additionally, my doctoral studies provided me with the opportunity to conduct a comprehensive review of the existing literature, which highlighted several methodological challenges, as detailed previously. Informed by these contextual and methodological insights, I conducted two studies. The first study aimed to explore and refine the methodology for measuring factors that contribute to students' general self-beliefs in mathematics. The second study aimed to investigate the significance of these factors within the context of mathematics instruction, specifically evaluating their role in predicting student engagement as an indicator of motivation.

According to the framework for motivational development presented in Chapter 2, student engagement is influenced by the teaching context. For instance, previous research suggests that students' cognitive engagement tends to increase during problem-solving and algorithmic thinking activities (Liu et al., 2023; Xing et al., 2023). These mathematical problems are similar to *unplugged* programming activities, which aim to enhance Computational Thinking (CT) in mathematics education, as discussed in Paper III. To explore this further, the second study was conducted in a mathematics course where students participated in contextually rich and challenging problem-solving tasks.

3.4 Measurement and operationalization

It is well known among scholars that theoretical concepts, such as students' beliefs about general mathematics self-efficacy, anxiety, and engagement, must be defined at an appropriate level of abstraction to make them observable (Kerlinger & Lee, 1999). This process, known as operationalization, is crucial in both quantitative and qualitative research. Despite the differing ontological and epistemological perspectives between methodologies such as participatory research and survey research, the core process of operationalization remains consistent. It involves aligning the research objectives in a manner that facilitates investigation (L. Cohen et al., 2018).

In the empirical studies presented in this thesis, the use of self-reported questionnaires for data collection aligns with the research objectives and may not require further justification. Nonetheless, it is important to acknowledge that all data collection methods, including self-reported questionnaires, are subject to inherent errors and challenges. Therefore, ongoing validation of self-report measures is essential. Generally, self-report instruments are effective for exploring student perceptions of motivation and engagement. This approach is particularly appropriate because motivation largely depends on individuals' perceptions of their environment and their roles within it (Greene, 2015). More specifically, there are five primary reasons supporting their use in investigating self-beliefs. Firstly, they are formulated in language familiar to respondents, facilitating meaningful interpretations. Secondly, individuals possess unique insights into their own personalities. Thirdly, people tend to invest greater effort when completing self-reports regarding their personality traits. Fourthly, the act of self-reporting engages an individual's sense of identity, thereby influencing their self-perception. Lastly, this method proves efficient due to its ease of reaching and addressing a large number of respondents (Paulhus & Vazire, 2007).

According to Michell (2012), a psychological attribute must possess a quantitative structure to be considered measurable. From this realist standpoint, it follows that most psychological attributes fail to meet this criterion and, as a result, are not considered measurable. For instance, consider two individuals who rate their self-confidence on a scale from 1 to 10: one selects a 2, while the other chooses 10. Determining whether one individual is five times more confident than another is challenging due to the subjective nature of self-confidence, suggesting that self-confidence may lack inherent quantifiability. Thus, from this realist perspective, the effective use of an interval scale requires a clear justification for how the degrees of an attribute can be ordered by magnitude.

Quantifiable attributes are advantageous in data analysis because they can be represented on interval scales, which are linear in nature and allow for the application of conventional statistical techniques (Stevens, 1946). However, while interval scales offer convenience in data analysis, they can also lead to potential drawbacks, such as the misinterpretation of categorical variables as though they were interval variables. For example, Likert-type scales, which are ranked categorical variables (ordinal variables), are often treated as though they were metric variables. Nevertheless, some scholars argue that this practice is justified, at least when the data approximates a normal distribution (Kerlinger & Lee, 1999), while others advocate for an even broader applicability (Norman, 2010). Another related area of debate involves the use of polychoric correlation, which is frequently recommended as an appropriate method for analyzing Likert-type scales (Chen & Popovich, 2002). However, research has shown that when Likert-type items contain more than six response categories, Pearson's correlation tends to yield coefficients nearly identical to those obtained using polychoric correlation (Rhemtulla et al., 2012). Nonetheless, Likert scales often fail to meet the criteria necessary to be considered truly metric across all datasets. Therefore, it is generally more appropriate to treat them as categorical variables (Rusch et al., 2017).

In conclusion, even when a psychological trait does not meet the criteria for a quantitative structure, it is important to recognize that a pragmatic approach to measurement does not necessitate quantifiability. Consequently, even if an attribute lacks a quantitative structure, objective facts can still be derived about that attribute. Furthermore, there is currently no method to verify whether a measurement accurately reflects the true value of the attribute being assessed (Kampen & Swyngedouw, 2000). In other words, despite an attribute's ontological subjectivity, it is still possible to achieve epistemic objectivity (Maul, 2013).

In line with the pragmatic approach advocated in this thesis, measurements can be defined as "processes aimed at producing information on properties of objects in the form of property values, and thus in the form of quantity values if the property is a quantity" (Mari et al., 2023a, p. 25). Furthermore, a measure is considered valid if the attribute being measured genuinely exists and if the causal relationship between variations in the attribute and the outcomes can be clearly explained (Borsboom et al., 2004). In light of this, the empirical studies presented in this thesis treat the observed variables as categorical while assuming that the underlying construct is metric and continuous (Kampen & Swyngedouw, 2000; Rusch et al., 2017).

3.4.1 Measures

In the surveys, participants were asked to indicate the frequency of certain feelings, beliefs, or behaviors related to their current mathematics class. For instance, they were prompted to report how often they believed they could perform well on a mathematics test. These items were designed to assess the strength of their sentiments regarding each statement.

A consistent five-point Likert-type scale was utilized, featuring the following response options: 1 = aldrig (*never*), 2 = sällan (*seldom*), 3 = ibland (*sometimes*), 4 = ofta (*often*), 5 = nästan jämnt (*usually*). This scale closely resembled the one used in the Mathematical Self-Efficacy and Anxiety Questionnaire (MSEAQ) by May (2009), except for the omission of the *no response* option. This omission was deliberate, as students were expected to evaluate statements regarding their mathematics self-beliefs or engagement. Additionally, the response option *usually* was translated as *nästan jämnt*, which, when back-translated word for word, means *almost always*. However, these response options differ from those in Skinner et al. (2009), whose engagement measure employed a four-point Likert-type scale ranging from 1 (not at all true) to 4 (very true). Although the measure resembles a frequency scale and might justify analyses typically associated with interval scales, it fundamentally categorizes responses into qualitative groups. This is evident in the subjective nature of distinguishing between terms such as *often* and *usually* when assessing one's confidence. Furthermore, given that these categories are ordered, the Likert scales employed are considered ordinal variables.

The original items in the MSEAQ were in English and were later translated into Swedish. Therefore, it was deemed essential to assess the reliability of these translated items. To achieve this, a back-translation process was employed. Additionally, Confirmatory Factor Analysis (CFA) was conducted to

Table 3.1. The operationalization of student engagement dimensions

Dimension	Description	Observational definition
Engagement/disengagement		
Behavioral	Engagement is characterized by active participation, including qualities like consistent effort, determination, focus, and attentiveness. Conversely, absence or deficiency of engagement is marked by a state of passivity or disinterest.	Engaged students demonstrate attentive listening and active involvement in tasks (e.g., "When I'm in this class, I listen very carefully"). Conversely, signs of passivity are evident when students merely pretend to work (e.g., "When I'm in this class, I just act like I'm working") (Skinner et al., 2008, 2009).
Emotional	Emotional states that fuel engagement include feelings of enjoyment, while disengagement may be characterized by sensations of frustration or boredom.	Students display emotions like interest, happiness, and enthusiasm while participating in activities (e.g., "When we work on something in this class, I feel interested"). Whilst disengagement induces feelings of disaffection, anxiety, or boredom (e.g., "When we work on something in class, I feel bored") (Skinner et al., 2008, 2009).
Cognitive	Described as employing profound learning strategies rather than superficial ones, as opposed to disorganized learning approaches.	Students demonstrate profound engagement by exploring multifaceted approaches to problem-solving or understanding mathematical concepts and procedures (e.g., "When I study for this class, I try to connect what I am learning with my own experiences"). In contrast to disorganized learning approaches (e.g., "I'm not sure how to study for this course") (H. Jang et al., 2016; Reeve, 2018).
Agentic	Defined as students' active involvement in the instructional process, involving continuous interaction and exchange between student and teacher.	Students actively engaging in asking questions, offering feedback to improve the learning environment for both themselves and their peers, and expressing their needs and preferences (e.g., "I let my teacher know what I need and want") (Reeve, 2013).

evaluate the fit indices. The results for the translated items of the MSEAQ indicated a satisfactory fit to the five-factor model proposed by May (2009), as detailed in Paper I. Similarly, the engagement items demonstrated a good fit to a five-factor model that includes the dimensions of student engagement: behavioral, emotional, agentic, cognitive engagement, and disengagement, as described in Paper IV.

Table 3.1 presents the dimensions and concepts employed in the empirical studies, along with their observational definitions. However, the dimensions and concepts related to general mathematics self-efficacy and anxiety are omitted from the table, as they have been investigated in the empirical studies (see Papers I, II, and IV). All the questionnaire items used in the empirical studies are presented in the appendix (Table A.2 and A.1).

General mathematics self-beliefs

Previous research has established various scales for measuring individuals' mathematics self-efficacy and anxiety beliefs (see e.g., May, 2009). For this study, the MSEAQ developed by May (2009) was chosen as the preferred instrument. This decision was based on the questionnaire's diverse range of item types, many of which are commonly used to assess general confidence in mathematics, such as "I believe I am the kind of person who is good at mathematics." Consequently, the MSEAQ includes items well-suited for EFA, particularly in examining the challenges associated with operationalizing general mathematics self-beliefs. Additionally, the MSEAQ was originally designed for college students aged 18-24, which makes it suitable for upper-secondary school students in Sweden, who are typically between 16 and 19 years old.

The MSEAQ consists of 29 items and, as noted by May (2009), addresses two distinct areas: mathematics self-efficacy and mathematics anxiety. It is important to clarify that, in this context, mathematics self-efficacy refers to a construct similar to self-concept. Specifically, constructs of this nature are commonly assessed through items such as "I believe I can learn well in a mathematics course" and "I believe I am the kind of person who is good at mathematics".

Student engagement

To assess student engagement, the empirical studies in this thesis utilize measures of behavioral, emotional, cognitive, and agentic engagement, as well as cognitive disengagement, as detailed in Table 3.1. Notably, only cognitive disengagement was selected from the original disengagement scales. This decision was based on the observed distinctiveness of cognitive disengagement from cognitive engagement in capturing different aspects of learning strategies, and concerns about redundancy among the other disengagement scales. However, one item from the original cognitive disengagement measure (H. Jang et al., 2016) was omitted due to challenges in translating the concept of a *study plan* into the Swedish school context. Nonetheless, it is important to note that the absence of disaffection or disengagement does not necessarily imply engagement (Jimerson et al., 2003). For instance, high scores on a Likert-type scale for engagement do not necessarily indicate that students are free from disengagement.

Concerns regarding redundant items were noted in several disengagement scales, which is significant because such redundancies can introduce excessively correlated variables, complicating data analysis. For example, certain items within the behavioral disengagement scale (Skinner et al., 2008) appear too similar to effectively distinguish between student responses. Specifically, the item "When I'm in class, I think about other things" has only subtle differences compared to "When I'm in this class, my mind wanders", potentially causing confusion among respondents. Similar issues were observed in the original agentic disengagement scale (H. Jang et al., 2016), particularly with the items "Most of the time in this class, I am passive" and "Most of the time in this class, I am silent and unresponsive". The redundancy issues discussed here mirror those explored in Papers I and II, which focus on general mathematics self-efficacy and anxiety.

Additionally, it is important to acknowledge that the constructs of emotional engagement (Skinner et al., 2008) bear considerable resemblance to those of intrinsic motivation (Ryan & Deci, 2020). Although the empirical research in this thesis does not directly assess intrinsic motivation, the overlap between these constructs is noteworthy. This overlap may complicate data analysis, particularly when examining or comparing these constructs simultaneously in correlation studies and across different studies.

3.5 Ethical considerations

Ethical considerations were taken into account in accordance with the ethical guidelines set forth by the Swedish Research Council (Vetenskapsrådet, 2017), and the principles outlined in the General Data Protection Regulation (GDPR[1]). Student participation in all studies was entirely voluntary, and written informed consent was obtained (see Figure A.1 in the appendix). Moreover, participants were informed of their option to withdraw their consent at any point, and the explicit purpose of data collection was clearly explained. Although questions about one's feelings are generally not considered sensitive personal information, measures were taken to safeguard participants' anonymity by refraining from collecting or sharing any personal details that could potentially identify individuals. Additionally, participants were asked to indicate their legal gender as either male or female.

As Hammersley and Traianou (2012) explain, ethical considerations are intertwined with the concept of research integrity and validity. Research integrity may pertain to the choice of research methodology. This selection is inherently tied to the research objective and is influenced by the researchers' epistemological perspectives (Hammersley & Traianou, 2012). For instance, qualitative researchers frequently choose interviews, whereas quantitative researchers lean towards surveys or experiments. Some researchers prioritize knowledge generation, while others emphasize practical or political aims such as participatory research. Given the context discussed, the central focus of this research is to identify the most suitable methodology for addressing situations characterized by substantial multicollinearity. Since this research centers on the analysis of quantitative data, no ethical concerns arose regarding the choice of data collection methods.

Research validity pertains to the credibility of findings in quantitative studies. A common misconception is that larger sample sizes automatically en-

[1] Source: https://www.imy.se/en/organisations/data-protection/

sure more reliable results (Shamoo & Resnik, 2015). However, larger samples do not necessarily lead to more reliable conclusions and may even result in exaggerated claims about the reliability of the findings (Hammersley & Traianou, 2012; Kline, 2020; Shamoo & Resnik, 2015; Wagenmakers, 2007). Therefore, it is crucial to balance the emphasis on the validity of results with a thoughtful consideration of their implications (Hammersley & Traianou, 2012; Shamoo & Resnik, 2015).

Throughout this thesis, I faced various subjective decisions inherent in quantitative research, especially concerning issues of validity. For example, these decisions include interpreting factor scores, visually representing data, and balancing statistical significance with practical implications. Additionally, they involve determining whether outliers reflect biases or genuine characteristics of the phenomena under study, as well as addressing missing values. Despite the guidance provided by the research community on these issues (see, for example, Tabachnick & Fidell, 2013), implementing these recommendations is not always straightforward, as it involves subjective judgment. Accordingly, ethical considerations were incorporated into each stage of the data preparation and statistical analysis processes.

For instance, rather than eliminating variables to address issues with missing values, I opted to impute the missing data. Additionally, instead of conforming data to fit the assumptions of traditional statistical tests, I did not remove outliers. Instead, I employed appropriate measures of central tendency and robust statistical methods to mitigate their impact, in line with the recommendations of Tabachnick and Fidell (2013) and Wilcox (2010). However, when there were indications that a reported value might be unreliable, such as an extreme response, it was excluded from subsequent analyses. To ensure transparency, I provided comprehensive explanations for the use of unconventional statistical methods, detailing both their rationale and implementation, drawing on insights from Shamoo and Resnik (2015) and Kline (2020). Furthermore, the intersection of ethics and research validity has shaped my perspective on generalization and the interpretation of findings, which I will further elucidate in the following section.

3.6 Generalization

As outlined earlier in this chapter, each sample is considered a distinct case shaped by its specific context. While conducting multiple case studies may enhance generalizability, there is often debate about the extent to which such studies can be generalized. Yin (2017) distinguishes between two types of generalization: analytic generalization and statistical generalization. Statistical generalization involves making inferences about a population based on the analysis of a representative sample. In contrast, analytic generalization seeks to extend findings beyond the specific case by treating the case as an example of a broader phenomenon. This approach helps researchers apply findings to other similar cases or contexts. Essentially, analytic generalization involves assessing whether findings from a case study can be generalized to other contexts by examining the influencing factors and settings. Thus, analytic generalizations operate at a more abstract conceptual level compared to those specific to individual cases (Yin, 2017).

In summary, consistent with my pragmatic stance, I generalize my findings through a combination of analytical generalization and statistical inference. Specifically, I treat each statistical sample as an individual case. Unlike Yin's definition of analytical generalization (Yin, 2017), I incorporate statistical inference to extend my findings to similar cases. Rather than relying solely on statistical generalization, I draw conclusions about similar cases within analogous contexts through the analysis of representative cases. This approach, therefore, facilitates a more nuanced interpretation of statistical generalization.

Several researchers (e.g., J. Cohen, 1994; Kline, 2020; Wagenmakers, 2007) have criticized the reliance on null hypothesis testing due to the frequent invalidity of the null hypothesis and the potential for misinterpreting p-values. It is important to emphasize that the p-value represents the probability of observing the data given that the null hypothesis is true, not the probability that the null hypothesis itself is true (J. Cohen, 1994; Wagenmakers, 2007). Furthermore, the use of p-values in data analysis carries several potential risks of misinterpretation (Kline, 2020; Wagenmakers, 2007). For instance, if two experiments—one small-scale and the other large-scale—yield identical p-values (indicating evidence against the null hypothesis), it might be mistakenly argued that the large-scale experiment provides stronger evidence against the null hypothesis (Wagenmakers, 2007). However, this is often not the case. In fact, the small-scale experiment might provide more robust evidence against the null hypothesis. This is because, when p-values are identical, the large-scale experiment typically has a smaller effect size due to its greater statistical power to detect differences (Wagenmakers, 2007).

In all statistical analyses conducted in the empirical studies, significance was assessed using the conventional threshold of $p < 0.05$. However, to mitigate the risk of misinterpreting p-values, interpretations were supplemented with confidence intervals, which offer a more comprehensive understanding of statistical significance. Additionally, I employed Bayesian analysis (see Paper IV), which evaluates the likelihood of both the null and alternative hypotheses simultaneously, thereby complementing frequentist methods. Furthermore, reflecting my pragmatic approach, each measurement is viewed as a social activity, emphasizing the research context in interpreting statistical significance within the frequentist framework. From this perspective, if the null hypothesis is deemed improbable, the findings suggest that other cases (samples) under similar conditions are also likely to exhibit improbability.

Data collection and analysis 4

The aim of this chapter is to present the data collection procedure and data analyses.

4.1 Overview

In both studies, the sampling procedure remained consistent. Initially, contact was made with the school principal via email, who then forwarded the message to the senior mathematics teacher. The senior teacher identified the teachers who were interested in participating in the study along with their students. Subsequently, three samples were collected on separate occasions. In each study, students were asked to complete an online questionnaire about their beliefs regarding their current mathematics class.

Figure 4.1 summarizes the data collection process. In Paper I, only the initial sample (Sample 1) was analyzed. Subsequently, Paper II explored the analysis of Samples 1 and 2, comprising 156 students. The final paper (Paper IV) focused solely on Samples 2 and 3. The decision to divide the research into Studies I and II was motivated by the overarching goal of examining the key factors behind students' general mathematics self-beliefs and their connection with student engagement. Additionally, combining both investigations within a single sample could potentially overwhelm participants with a large number of items to evaluate, thereby increasing the risk of generating low-quality responses (Denscombe, 2010).

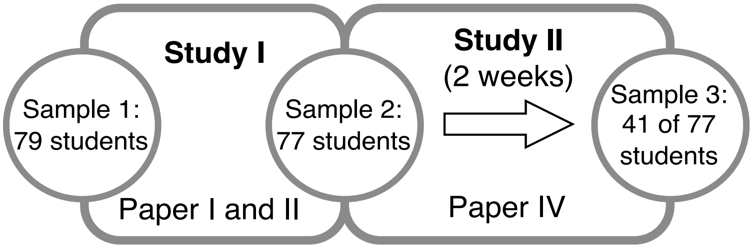

Figure 4.1. Summary of data collection

4.2 School context and participants

All participating students were enrolled in an upper-secondary school-level university entrance qualification program. In Sweden, upper secondary school spans from year 10 to year 12, with students typically aged 15 to 19.

The first sample comprised 79 upper-secondary students, 73% of whom were female and aged 16-17. These students were selected from a school with approximately 574 students, of whom 67% were female. Among the students at this school, 383 were aged 16-17. In Sweden, a total of 70 791 students were enrolled in the university entrance qualification program offered at the experimental school, with 58% of them being female. The second sample comprised 77 upper-secondary students (38% female), aged 18–19, who were enrolled in their final semester of upper-secondary school. These students were taking a mathematics course known for its advanced theoretical and mathematical content (Mathematics 5), which emphasizes problem-solving with complex, real-world scenarios as a national objective. Additionally, they were part of a specialized upper-secondary program focused on technology and science. Of the total 236 students enrolled in this program, 91 were in the same age group. Across Swedish upper-secondary schools, approximately 21 625 students were enrolled in similar programs, with 21% of them being female.

In the second and third samples, the same group of students completed identical questionnaires separated by a two-week interval. However, in the third sample, 41 of the 77 students from the second sample participated, with females comprising 34% of this group.

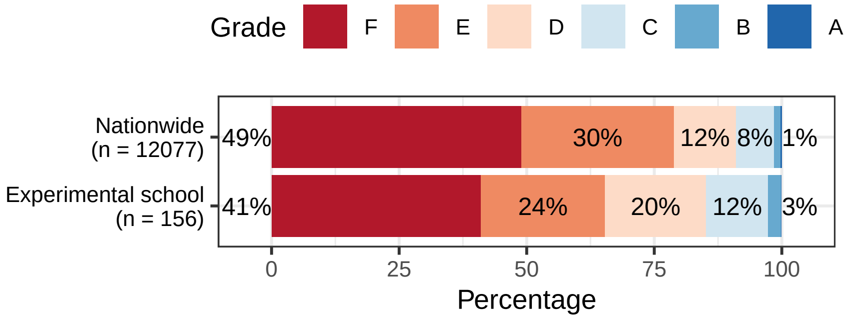

Figure 4.2. A summary of the national test results for Mathematics 2b from schools nationwide, including the experimental school, as reported by the Swedish National Agency for Education in 2019. The grading scale ranges from F (fail) to A (the highest grade)

4.2.1 Representativeness

To assess the relevance of the findings from the experimental school to similar educational settings, I examined the similarity between the distribution of national test results at the experimental school and the national distribution. For this analysis, I utilized test results from a compulsory basic mathematics course (Mathematics 2b) in Sweden, collected in 2019. Figure 4.2 illustrates that the grade distribution at the experimental school is largely consistent with the national distribution, showing only minor deviations. To complement this visual assessment, additional statistical analyses were conducted.

A conventional chi-square test indicates a statistically significant difference between the outcomes of the experimental school and the nationwide results, $\chi^2(5) = 17.86, p = 0.0031$. However, given that the chi-square test is highly sensitive to sample size, even minor differences may appear statistically significant (Jamil et al., 2017). Therefore, this result raises concerns about the true significance of the findings. To address this issue, a Bayesian analysis was conducted using the Bayes Factor (BF), which represents the odds ratio of the likelihood of the alternative hypothesis relative to the null hypothesis (BF_{10}) or vice versa (BF_{01}). Unlike traditional null hypothesis testing, Bayesian analysis evaluates the likelihood of both the null and alternative hypotheses simultaneously (Jamil et al., 2017). The function `contingencyTableBF` in R-package *BayesFactor* (Version 0.9.12.4.4; Morey & Rouder, 2022) yielded a

Bayes factor of $BF_{01} = 429\,582$, indicating extremely strong evidence in favor of the null hypothesis. Consequently, the Bayesian analysis provides substantial evidence supporting the absence of differences in national test grades for the Mathematics 2b course between the nationwide results and those of the experimental school. This suggests that, in terms of mathematics achievement, the experimental school is representative of a typical upper-secondary school in Sweden.

4.3 Study I

The objective of this study was to enhance the methodology for utilizing EFA in the analysis of students' general mathematics self-beliefs and to investigate the factors identified through this approach. The research was conducted in two phases. The initial phase focused on assessing the effects of excluding certain items to mitigate issues related to extreme multicollinearity within the data. During this phase, self-reported data concerning students' general mathematics self-beliefs were gathered using the items listed in Table A.1 in the appendix. The results of this phase are detailed in Paper I. Building on these findings, the second phase involved selecting 17 items (see Table A.3 in the appendix) for a more in-depth analysis to uncover the underlying factors of students' general mathematics self-beliefs. The outcomes of this examination are presented in Paper II.

4.4 Study II

Building on the findings from Study I, this research aimed to evaluate student engagement and their general beliefs about mathematics. To achieve this, the study employed the instruments detailed in Tables A.2 and A.3 in the appendix. Prior research indicates that individuals' beliefs about their mathematical abilities may shift depending on the complexity of the tasks they face (Street et al., 2022). Furthermore, engaging in problem-solving activities designed to enhance algorithmic thinking has been shown to boost cognitive engagement (Liu et al., 2023; Xing et al., 2023). To account for potential variations in student responses throughout the mathematics course segment, these measures were administered twice to the same group of students: once at the beginning and once at the end of the segment. During this segment, students participated in collaborative problem-solving involving both mathematically challenging and contextually realistic scenarios. This approach demonstrates one method of integrating programming into mathematics education. For a detailed discussion of different interpretations of programming within mathematics education, see Paper III.

Initially, students were organized into problem-solving groups as part of the study, which was conducted within the framework of conventional mathematics instruction. The classroom teacher managed all aspects of classroom activities, including the formation of these problem-solving groups. Students were then introduced to the problem-solving elements of CT, which encompass techniques such as deconstructing problems into smaller, manageable parts, identifying patterns and similarities, employing abstraction, devising solutions in a sequential manner, verifying the accuracy of solutions, and selecting the most efficient and generalizable strategies (see ISTE, 2011; Selby & Woollard, 2014). To practice these techniques, students engaged with the following introductory problem designed to apply these principles:

> You are faced with a challenge. Given two containers, one with a capacity of 4 milliliters and the other with a capacity of 7 milliliters, how can you measure exactly 5 milliliters using these containers? Present your solution in a table format, aiming for the most efficient solution with the fewest steps. (Adapted from Australian Government Department of Education, Skills, n.d.)

After addressing the initial problem, students in each group selected a specific problem to focus on for the subsequent two weeks. Some groups chose to tackle their problem within the mathematics classroom, while others preferred to move to a nearby room to collaborate without disturbance. The problems selected by the students were freely chosen and were readily available in a specific chapter of their standard mathematics textbook (see Alfredsson et al., 2019). For instance, one of the problems featured in the textbook was a variation of the classic Jeep Problem, also known as the Desert Crossing Problem (Alfredsson et al., 2019, p. 243). Originating in 1947 (Fine, 1947), this logistical problem involves a jeep starting from a base camp with unlimited access to fuel but limited carrying capacity. The jeep can transport and store fuel at depots along its route to refuel while away from the base camp. Students were tasked with calculating the minimum amount of fuel required to traverse the desert.

4.5 Data analyses

As the specific technical aspects of each study have been thoroughly discussed in Papers I, II, and IV, they will not be reiterated here. Instead, this section will focus on a more thorough explanation of the rationale behind certain methodological choices, with particular emphasis on data preparation and the analysis of latent variables.

4.5.1 Data preparation and software selection

All data processing and analyses were conducted using the statistical software R (Version 4.3.1; R Core Team, 2023). This software was selected for several reasons. Firstly, it eliminates the need for spreadsheet programs in data cleaning and dataset merging. Secondly, R offers a wide range of statistical packages, including modern methods that are not easily accessible in other statistical software such as IBM SPSS Statistics. Moreover, R is open-source, making it freely available, and it supports literate programming (see definition by Knuth, 1984), which allows for the simultaneous manipulation of text and code. For example, in the creation of this thesis, I used R (Version 4.3.1; R Core Team, 2023) and the R-packages *bookdown* (Version 0.37; Y. Xie, 2016), and *rmarkdown* (Version 2.23; Y. Xie et al., 2020, 2018), which facilitated the integration of R's capabilities directly within the Markdown language.

In the data preparation process, the data underwent multiple stages of processing and analysis. All analyses began with data cleaning to eliminate errors and anomalies, ensuring that the data were suitable for further examination. However, outliers were retained unless they were suspected to be erroneous responses. The rationale for this decision is detailed in the preceding section on ethical considerations.

Furthermore, following the recommendation by Brown (2015), missing values were imputed using multiple imputations with the Multinomial Logit model, utilizing the R package *mice* (Version 3.16.0; van Buuren & Groothuis-Oudshoorn, 2011). Five imputations were performed, each iterated 50 times. For imputation, 15 to 25 predictor variables were considered (van Buuren & Groothuis-Oudshoorn, 2011). Missing values for the mathematics self-belief variables were imputed using other self-belief variables as predictors, following the same approach as for the engagement variables. A critical assumption for the application of multiple imputation techniques is that the data are Missing Completely at Random (MCAR). Therefore, this assumption was evaluated through a statistical test using the R package *naniar* (Version 1.0.0; Tierney & Cook, 2023). Table A.4 in the appendix summarizes the test results, indicating that the missing values in all samples (1, 2, and 3) are likely to be MCAR.

4.5.2 Overview

Following the data preparation phase, polychoric correlations were computed to identify variables exhibiting significant intercorrelations. This evaluation involved both qualitative assessments and quantitative criteria. Additionally, careful consideration was given to the potential consequences of excluding variables with very high correlations to address concerns about extreme multicollinearity. The results of these analyses constitute the central findings presented in Paper I. Based on these insights, 17 items were selected for inclusion in a questionnaire administered in Study II. This questionnaire was analyzed using a EFA with polychoric correlations. To determine the primary factor contributing to gender differences, regression factor scores were computed, followed by a series of statistical tests, including MANOVA, ANOVA, and logistic mixed models. The results of this analysis are detailed in Paper II. Building on these findings, Paper IV investigated the predictive effects of the identified factors (outlined in Paper II) on student engagement through a series of mixed model analyses. These analyses employed factor scores derived from Item Response Theory (IRT) models.

The following sections will examine two analyses: identifying factor structure using EFA with polychoric correlations, and predicting student engagement using IRT model factor scores.

4.5.3 Identifying factor structure: EFA with polychoric correlations

In this thesis, EFA was employed to uncover the latent dimensions underlying variables related to general mathematics self-beliefs. The choice of EFA over CFA was primarily motivated by the exploratory nature of the study, which aimed to generate new insights rather than confirm pre-established theoretical models. Additionally, given the complexities in defining and measuring general mathematics self-efficacy beliefs, EFA was deemed the most appropriate method. Principal Component Analysis (PCA) was considered unsuitable for this context, as it focuses on capturing linear variation in the data, similar to discriminant analysis, and deviates from the primary objective of estimating underlying factors (Field et al., 2013). Therefore, EFA was selected as the best-suited approach for this analysis.

Numerous sources offer guidance on performing EFA. In this thesis, I reviewed a range of these references (Beavers et al., 2013; Costello & Osborne, 2005; Yong & Pearce, 2013), in addition to consulting relevant statistical textbooks (Field et al., 2013; Tabachnick & Fidell, 2013). Despite the availability of guidelines, there is still a notable absence of comprehensive criteria for determining the suitability of a correlation matrix for factor analysis, particularly in the presence of substantial multicollinearity. Many existing guidelines tend to rely on informal rules or heuristics. A common heuristic suggests that variables with high intercorrelations, often defined as those with correlation coefficients exceeding 0.80, should be addressed by removing one of the highly correlated variables (L. Cohen et al., 2018; Field et al., 2013). Another widely used heuristic for evaluating multicollinearity involves calculating the Squared Multiple Correlation (SMC) for each variable (Tabachnick & Fidell, 2013). A high SMC value suggests that a variable is strongly correlated with other variables in the dataset, indicating significant multicollinearity. In such cases, it is advisable to consider excluding the variable from the analysis. However, a notable challenge is the lack of a well-defined threshold in the literature for determining what constitutes an excessively high SMC value. Nonetheless, the estimation of the SMC is based on the determinant of the correlation matrix. A determinant value approaching zero signifies a high degree of multicollinearity (Tabachnick & Fidell, 2013). Given that this approach relies on heuristic methods, it is prudent to ensure that the determinant of the correlation matrix remains above a very low threshold, specifically set close to zero. Therefore, in the empirical studies, a heuristic threshold of 0.00001 was applied (Field et al., 2013).

Two main approaches to EFA exist: one through the Classical Test Theory (CTT) framework and the other through the IRT framework. Typically, CTT posits a linear relationship between latent variables and observed scores, whereas IRT assumes a non-linear relationship (Rusch et al., 2017). In the context of the CTT framework, a notable limitation arises when latent variables are treated as metric, which implies that observed scores are also regarded as metric. However, Likert scales often do not consistently satisfy the criteria for being considered metric (Rusch et al., 2017). This inconsistency has sparked ongoing debates among researchers about whether to employ IRT or CTT frameworks (Robitzsch, 2020). A specific point of debate involves the application of Pearson's correlations to Likert-type scales, which are sometimes assumed to be approximately metric (Norman, 2010). However, this practice is among the most frequently misused approaches in social and behavioral research (Chen & Popovich, 2002), primarily because Pearson's correlations are sensitive to non-normally distributed data and, in particular, to outliers (Wilcox, 2012).

Despite the advantages of the linear assumption in CTT, such as the simplicity and conventionality of statistical methods, this assumption presents several methodological challenges. One such challenge is the difficulty in making accurate claims about the latent trait when using sum scores, which is a common practice under the assumption of a continuous latent trait (Rusch et al., 2017). To address these limitations, I adopted an approach similar to the IRT framework (Wirth & Edwards, 2007), estimating polychoric correlations to preserve the ordinal nature of the observed variables rather than treating them as metric and continuous (Chen & Popovich, 2002; Choi et al., 2010). This correlation method was employed in Papers I and II. Mapping ordinal variables onto a theoretical underlying continuum is commonly referred to as the underlying variable approach (Kampen & Swyngedouw, 2000). However, the validity of the estimated correlation coefficient depends on the extent of evidence supporting the assumption of underlying bivariate normal latent variables. In line with the suggestion of Kampen and Weeren (2017), I conducted statistical tests for bivariate normality on the observed variables (see Tables A.5 and A.6 in the appendix). The bivariate chi-square test results were obtained using the `polychor()` function from the R package *polycor* (Version 0.8.1; Fox, 2022). Although some bivariate tests indicate that bivariate non-normality cannot be entirely ruled out, as suggested by the statistical significance in chi-square tests, the overall findings support the use of polychoric correlation.

4.5.4 Predicting student engagement: Mixed models with IRT model factor scores

The purpose of Paper IV is to examine how general self-beliefs in mathematics predict student engagement. To achieve this, estimating factor scores is crucial. Given the assumptions that the latent variables follow a normal distribution and that there is a non-linear relationship with the observed variables—consistent with IRT (Rusch et al., 2017)—one method for estimating these factor scores is through CFA using polychoric correlations. This can be accomplished using the `WLSMV` estimation method with the `cfa()` function in the R package *lavaan* (Version 0.6.17; Rosseel, 2012). However, reliable estimates could not be obtained with very small sample sizes, such as the sample of 41 observations in Sample 3. Therefore, a Bayesian approach was utilized. The Bayesian framework is particularly suited for small sample sizes (McNeish, 2016; Mioevi, 2020) and offers a more robust alternative to traditional statistical significance testing, which can be prone to misinterpretation within the CTT framework (Kline, 2020).

In the Bayesian framework, all parameters are considered random variables, whereas in the CTT framework, it is the data that are treated as random (McNeish, 2016). This distinction means that parameter estimates in Bayesian analysis inherently possess a probabilistic interpretation. Bayesian inference is grounded in Bayes' theorem, which updates prior beliefs about the parameters with observed data to form posterior beliefs. The posterior distribution obtained provides a probabilistic summary of the parameters, allowing for estimates of their values within specific ranges (Mioevi, 2020). However, estimating the posterior distribution can be problematic with small sample sizes. To address this issue, I employed Markov Chain Monte Carlo (MCMC)[1] methods, as suggested by e.g., Mioevi (2020), to approximate the posterior distribution. Specifically, I estimated the factor scores from the posterior distributions using the R package *mirt* (Version 1.41; R. P. Chalmers, 2012).

In both Paper II and Paper IV, linear mixed models were employed to predict specific outcomes. However, before conducting the analyses, a careful assessment was made to determine the suitability of mixed-effects models over fixed-effects models (see Paper II). In Paper II, students' course levels were treated as random effects to account for the variability in mathematics content across different courses, which could influence their mathematics self-beliefs. In the subsequent study presented in Paper IV, the data comprised repeated measurements, resulting in dependent observations and residuals (Field et al., 2013). To address this, a mixed linear model was applied to estimate the relative contributions of each predictor variable. This approach avoids the sphericity assumption required by repeated-measures ANOVA while still allowing the interpretation of the model as a linear model (Field et al., 2013). Furthermore, the mixed linear model facilitates the detection of longitudinal

[1] Wirth and Edwards (2007) describes Monte Carlo integration using the example of estimating the area of a circle. The method involves enclosing the circle within a square, randomly placing points inside the square, and then estimating the area of the circle by scaling the ratio of points that fall inside the circle relative to the total number of points, with this ratio multiplied by the area of the square. As the number of points increases, the estimation becomes more accurate. This technique is also applicable to complex geometric shapes.

differences by incorporating random effects to account for the correlations among students' repeated responses.

In general, similar to linear regression analysis, several key assumptions must be satisfied to ensure the generalizability of results from linear mixed models. These assumptions, as outlined by Berry (1993), include homoscedasticity (the requirement for equal variance of residuals across levels of the predictor variables), the absence of outliers, the independence and normal distribution of errors, and the absence of perfect multicollinearity among predictor variables. However, in educational research, it is not uncommon for data to violate one or more of these assumptions. Consequently, in Paper II, I examined factor score distributions for signs of non-normality, such as skewness and outliers. Significant evidence of non-normality was identified in the data analysis of Paper II. Therefore, both conventional linear mixed models using the R package *lme4* (Version 1.1.34; Bates et al., 2015), and robust linear mixed models using the R package *robustlmm* (Version 3.2.0; Koller, 2016) were employed to address these issues.

The results obtained from linear mixed models were supplemented by Bayesian estimates (Yamamoto & Miyazaki, 2024), which were derived using the R package *brms* (Version 2.21.0; Bürkner, 2018, 2021, 2017) and fitted with a Bayesian generalized multivariate model using MCMC. However, smaller sample sizes can increase the influence of prior information on parameters, potentially introducing biases, as extreme values may disproportionately affect results compared to larger samples (McNeish, 2016). In Paper IV, due to the lack of readily available prior information, less informative priors were used. This decision was motivated by the desire to let the data speak for itself and to avoid introducing potential bias from overly informative priors (McNeish, 2016; Yamamoto & Miyazaki, 2024; Zondervan-Zwijnenburg et al., 2017). The Bayesian approach with MCMC allowed for the estimation of both the direct and indirect contributions of each variable. To address potential biases, the Bayesian direct effect estimates were cross-verified with frequentist estimates. The results presented in Paper IV demonstrate that both Bayesian and frequentist estimates were comparable, thus confirming their reliability.

5 Synthesis of Findings

The primary focus of this chapter is to synthesize the findings in relation to the overarching research questions of this thesis. I will refrain from extensively restating the results of each individual paper. For a more thorough examination of these results, please refer to Papers I-IV. However, I will discuss supplementary findings related to gender disparity, which were not explicitly detailed in Paper II. The research contributions are presented and organized in alignment with the two aspects of this thesis, as outlined in Section 1.2.

5.1 Operationalization challenges

The key findings and contributions of my research, which address methodological challenges associated with validity and extreme multicollinearity in the context of EFA, can be summarized as follows:

- Previously validated instruments for measuring general mathematics self-beliefs need reevaluation.
- In EFA, relying solely on quantitative criteria may overlook important information about latent factors. Therefore, to avoid excluding valuable variables, it is crucial to employ qualitative evaluation.
- Addressing the issue of multicollinearity in EFA is crucial for overcoming challenges related to the operationalization of constructs, particularly in distinguishing between *generalized mathematics self-efficacy* and *mathematics self-concept*.
- A four-factor model describes students' general self-beliefs in mathematics, with *in-class anxiety* emerging as the most significant factor in explaining gender differences.

- The alignment between male students' *mathematics self-concept* and their prior mathematics achievement is weaker compared to that observed in female students.

The methodological challenges in investigating general mathematics self-efficacy beliefs are examined in Papers I and II. These challenges concern validity and extreme multicollinearity that may arise during EFA. In Paper I, we investigated the potential loss of information resulting from the selective removal of problematic variables in EFA, without conducting a qualitative analysis of the omitted variables. Additionally, Paper II identifies and analyzes the latent factors derived from EFA after addressing issues with validity and extreme multicollinearity.

In the following sections, I will provide a detailed discussion of my research contributions. These contributions are organized into two main areas: the use of qualitative reasoning within EFA and the rationale for choosing EFA over CFA. Additionally, I will present supplementary findings related to gender differences.

5.1.1 Qualitative reasoning in EFA

Extreme multicollinearity is typically managed by increasing the sample size, eliminating one or more highly correlated variables based on quantitative criteria, or combining the multicollinear variables (Tabachnick & Fidell, 2013). However, in educational research, it is often undesirable to combine multicollinear variables and impractical to increase the sample size. Therefore, the most feasible solution is usually to eliminate one or both of the problematic variables. Nevertheless, as demonstrated in Paper I, relying exclusively on quantitative criteria is insufficient, as it can lead to an oversimplification of information regarding latent variables. This conclusion is also supported by the inconsistency in the recommended quantitative thresholds for omitting variables found in the literature. While some researchers suggest excluding variables only when the intercorrelations are exceedingly high, specifically above 0.90 (Tabachnick & Fidell, 2013), others advocate for a more cautious approach, recommending further examination when intercorrelations exceed

0.60 (L. Cohen et al., 2018; Rockwell, 1975). Nonetheless, as Rockwell (1975) argues, "coping with multicollinearity is not simply a matter of locating and removing highly interdependent variables" (p. 318), especially since these variables are theoretically grounded. In other words, intercorrelations that may not appear problematic according to specific quantitative criteria can, in fact, be the root cause of significant multicollinearity (Rockwell, 1975).

In conclusion, extensive multicollinearity, regardless of sample size, can be effectively addressed by removing one or more of the problematic variables. However, this decision must be guided by both qualitative reasoning and quantitative correlation thresholds. Qualitative reasoning, in particular, involves assessing highly intercorrelated variables for potential redundancy or low validity when severe multicollinearity is present (see Paper I). For example, following qualitative analysis, the item "I worry that I will not be able to get an A in my mathematics course" was deemed to have low validity in the research context and was therefore excluded. This decision was based on its lack of correlation with responses to the statement, "I believe I can get an A in a mathematics course." Additionally, concerns were raised regarding the validity of this item, as students might focus on achieving a grade they consider satisfactory even if it does not meet the A level[1]. Retaining this item could thus potentially compromise the interpretation of the results.

5.1.2 The significance of EFA

Considering the existence of previously validated instruments for measuring general mathematics self-efficacy and anxiety beliefs, the most pertinent question emerges: why should EFA be conducted in the first place instead of CFA, and why is it essential to address issues with multicollinearity?

As previously discussed, the literature identifies significant challenges in effectively implementing measures for assessing general mathematics self-beliefs (Bong & Skaalvik, 2003; Klee et al., 2022; Lee et al., 2020; Marsh et al., 2019). These challenges are evident across various validated questionnaires. This thesis specifically examines the MSEAQ, which includes items

[1]In Sweden, students are graded on a scale ranging from F (fail) to A (the highest grade).

from several established questionnaires (see May, 2009), all of which have been validated through large-scale studies.

Nevertheless, large-scale studies often encounter challenges due to ambiguous operational definitions. For example, Marsh et al. (2019) identified difficulties in empirically differentiating between *generalized mathematics self-efficacy* and *mathematics self-concept*. These difficulties may stem from issues with unclear operationalization. Specifically, the item "I am convinced that I can understand even the most difficult content in math" was used to measure *generalized mathematics self-efficacy*, while "It is easy to understand things in math" was intended to assess *mathematics self-concept*. Such subtle differences in wording can potentially confuse respondents and affect the validity of the study's conclusions. Similarly, the findings from Paper II suggest that the item "I believe I can understand the content in a mathematics course" pertains to the construct of *generalized mathematics self-efficacy*. However, it is empirically separate from the item "I believe I am the kind of person who is good at mathematics", which pertains to the construct of *mathematics self-concept*. This distinction emphasizes that *generalized mathematics self-efficacy* specifically concerns understanding content within a mathematics course, whereas *mathematics self-concept* reflects a broader self-image related to overall proficiency in mathematics.

Furthermore, it is practically useful to distinguish between *generalized mathematics self-efficacy* and *mathematics self-concept*. Generalized mathematics self-efficacy beliefs are extensively studied in educational research and are often used to predict outcomes such as academic achievement. However, the effectiveness of these predictions depends on the level of specificity involved (Bong & Skaalvik, 2003; Pajares, 1996). General beliefs about one's abilities tend to be more predictive of overall grades, while more specific beliefs about particular tasks or content are better at predicting performance in those specific areas. Consequently, understanding these distinctions is crucial, as they can significantly impact the interpretation of research findings and have implications for various areas of educational research, including studies on gender differences.

For example, research has indicated that items on scales assessing mathematics self-concept (e.g., "I am not just good at mathematics") and mathematics self-efficacy (e.g., "How confident are you in solving an equation?") in the PISA 2003 and 2012 assessments may lack reliability for cross-country comparisons (Ding et al., 2023). The challenge of comparing results across different countries may arise from variations in self-reporting styles, shaped by cultural influences (Paulhus & Vazire, 2007). To address and potentially mitigate these challenges, this thesis proposes an approach that interprets findings within the unique context of each research setting. Furthermore, a comprehensive analysis of the observed variables is crucial for identifying and addressing validity concerns, as well as minimizing problematic intercorrelations.

The factor structure

Through a careful process of selectively excluding observed variables, guided by qualitative and quantitative assessments prior to performing EFA, it was possible to empirically distinguish between closely related constructs. The findings from Paper II revealed four significant latent factors that explained students' general mathematics self-beliefs. Two of these factors were identified as self-concept-like (Marsh et al., 2019), reflecting students' *mathematics self-concept* and *generalized mathematics self-efficacy*. This finding is significant due to the inherent methodological challenges in operationalizing these constructs in a manner that ensures clear delineation.

The factor model comprised two factors reflecting students' anxiety or worry in class. One factor represented students' anxiety about asking questions in class (labeled as *in-class anxiety*), while the other factor represented students' anxiety related to completing assignments and engaging with mathematics at home (labeled as *assignment anxiety*). The self-concept-like factors were found to exhibit strong associations with each other, as well as with *assignment anxiety*.

In-class anxiety shares similarities with the concept of *fear of failure* (see A. J. Martin & Marsh, 2003); however, important distinctions exist between the two. Conceptually, *fear of failure* can be divided into two orientations: individuals who seek to avoid failure by pursuing success, and those who aim to avoid the negative consequences associated with failure. *Fear of failure* is typically measured using statements such as "I often do not even try because I am afraid of making mistakes", where respondents explicitly associate their fear with a specific cause. In contrast, *in-class anxiety*, as reflected in items like "I am afraid to give an incorrect answer during my mathematics class", involves respondents' fear of providing wrong answers and feeling nervous when participating in classroom interactions, during mathematics. This form of anxiety notably lacks an explicit cause and is more narrowly focused on the act of learning and knowledge demonstration. Consequently, *in-class anxiety* has a stronger epistemological dimension compared to *fear of failure*.

In the existing literature, the only comparable factor model is reported in the study by May (2009). This research employed factor analysis to reveal a five-factor structure, with the factors being General Mathematics Self-Efficacy, Grade Anxiety, Future, In-Class, and Assignment. Notably, the Grade Anxiety and Future factors included items that assessed both general mathematics self-efficacy beliefs and mathematics anxiety. Unlike the findings discussed in Paper II, the factor structure identified by May (2009) presents greater challenges in interpretation. For example, the General Mathematics Self-Efficacy factor encompasses items that evaluate both an individual's *mathematics self-concept*, which is past-oriented, and their future-oriented perceptions of *generalized mathematics self-efficacy*. This includes statements such as "I believe I am the kind of person who is good at mathematics", which reflects *mathematics self-concept*, and "I believe I can do the mathematics in a mathematics course", which pertains to *generalized mathematics self-efficacy*. Additionally, the item "I believe I will be able to use mathematics in my future career when needed" was found to load on both the General Mathematics Self-Efficacy factor and the Future factor. This observation indicates a significant overlap among the variables, which may lead to ambiguous factor interpretations. Furthermore, the ambiguous factor structure observed may be partly attributed to substantial multicollinearity within the data. These issues associated with multicollinearity have been examined and addressed in Papers I and II.

In Papers I and II, the challenges associated with factor interpretation are addressed through a meticulous process of item selection and removal, specifically targeting items with high correlations. For instance, in Paper II, the item "I worry that I will not be able to complete every assignment in a mathematics course" was distinctly associated with the *assignment anxiety* factor. This approach diverges from study by May (2009), who retained this item despite its significant cross-loading onto the future factor. Despite differences in the theoretical frameworks, there are significant similarities between the five-factor model presented by May (2009) and the four-factor models discussed in Paper II. Notably, the conceptualization of the in-class factor and the assignment factor aligns closely across these models.

Furthermore, the results from Paper II suggest that, despite no observed gender differences in prior mathematics achievement, indications of gender disparities emerged across all factors. Logistic mixed-effects models, after controlling for all four latent factors, revealed that only *in-class anxiety* remained statistically significant. Furthermore, these models suggest that female students have a more accurate perception of their mathematical competence, as evidenced by their higher likelihood of achieving a grade A in previous mathematics courses compared to a grade C. In contrast, male students showed a higher likelihood of achieving a grade E. One possible explanation for these gender differences in *in-class anxiety* may be attributed to the influence of gender stereotypes, such as the belief that boys are better at mathematics than girls (Goetz et al., 2013).

5.1.3 Supplementary findings

In Chapter 4, I elaborate on the rationale behind the selection of statistical methods employed. The Bayesian approach is particularly noteworthy, as it is well-suited for addressing the complexities inherent in educational research. One of the strengths of Bayesian analysis is that it does not typically require large datasets to generate meaningful inferences, although the quality of these inferences generally improves with larger sample sizes (McNeish, 2016). Moreover, while large sample sizes may provide more stable estimates in statistical analysis, it is crucial to remain mindful of the ethical considera-

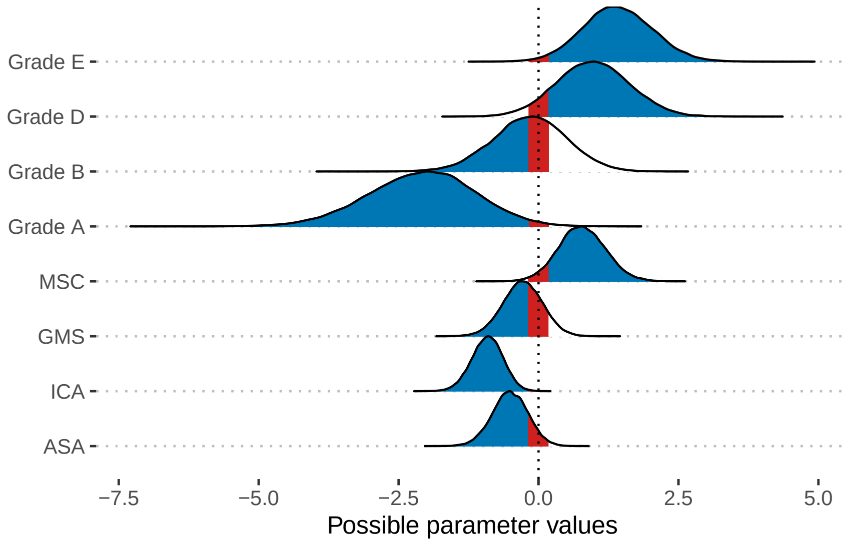

Figure 5.1. The posterior distributions for the predictors of gender including ROPE range, where MSC = Mathematics self-concept, GMS = Generalized mathematics self-efficacy, ICA = In-class anxiety, ASA = Assignment anxiety

tions involved. For example, statistical significance can be inflated in large-scale studies, potentially overstating the importance of the results (Hammersley & Traianou, 2012; Kline, 2020; Shamoo & Resnik, 2015). In conclusion, the Bayesian framework provides a practical approach to data analysis, particularly in educational research, where data often deviate from ideal assumptions of randomness. Furthermore, Bayesian analysis with MCMC can address sample size limitations by making the quality of inference less dependent on the amount of data. This approach was utilized in Paper IV.

To further investigate gender differences beyond the analysis presented in Paper II, Bayesian generalized multilevel models were employed to derive the posterior distribution for each factor (Figure 5.1). These models are based on the factor scores established in Paper II and were estimated using MCMC methods, as implemented in the R package *brms* (Version 2.21.0; Bürkner, 2018, 2021, 2017). In these models, course level was included as a random effect. Table 5.1 presents the results of the Bayesian significance tests con-

Table 5.1. Bayesian fit indices

Parameter	Median	pd	% in ROPE	Rhat
Intercept	$-0.88[-2.90, 1.04]$	86.53	8.28	1.001
Grade E	$1.40[0.22, 2.65]$	99.00	1.62	1.000
Grade D	$0.97[-0.29, 2.26]$	93.55	7.41	1.000
Grade B	$-0.15[-1.48, 1.12]$	59.17	21.39	1.000
Grade A	$-2.06[-4.04, -0.29]$	98.89	1.23	1.000
MSC	$0.77[-0.03, 1.62]$	97.02	6.37	1.000
GMS	$-0.28[-0.99, 0.41]$	78.84	29.37	1.000
ICA	$-0.91[-1.48, -0.39]$	99.98	0.33	1.000
ASA	$-0.51[-1.14, 0.11]$	94.81	13.77	1.000

Note: $N = 154$. 95 % Credible Intervals (CI). ROPE 100 % CI $= [-0.18, 0.18]$. Prior distribution (weakly informative) $= Normal(0, 10)$. MSC = Mathematics self-concept, GMS = Generalized mathematics self-efficacy, ICA = In-class anxiety, ASA = Assignment anxiety

ducted on the posterior distributions (with 20.000 iterations), including the median and 95% Credible Intervals (CI).[2].

The posterior indices and the region of Region of Practical Equivalence (ROPE)-based indices were computed using the R package *bayestestR* (Version 0.13.2; Makowski, Ben-Shachar, & Lüdecke, 2019). The probability of direction (pd) represents the likelihood that a parameter is strictly positive or negative. This pd value corresponds to p-value thresholds commonly used in a frequentist framework: pd values of 95%, 97.5%, 99.5%, and 99.95% align with p-values of 0.1, 0.05, 0.01, and 0.001, respectively (Makowski, Ben-Shachar, Chen, & Lüdecke, 2019). The ROPE defines a range within which parameter values are considered negligible or of limited practical significance. The proportion of the posterior distribution falling within the ROPE, as presented in Table 5.1, follows the guidelines established by Makowski, Ben-Shachar, Chen, and Lüdecke (2019) and is used to assess the significance of the results. Specifically, if less than 2.5% of the posterior distribution lies within the ROPE, the effect is considered statistically significant, indicating a rejection of the null hypothesis. Conversely, if more than 97.5% of the distribution is

[2] Credible intervals represent a specified percentage of the posterior distribution, indicating, in this case, a 95% certainty that the true value of the median lies within the given range.

within the ROPE, the effect is deemed negligible. For values falling between these thresholds, the effect remains inconclusive (Kruschke, 2018).

In conclusion, the influence of grade E on gender shows a 99% probability of being positive, indicating a statistically significant and substantial effect (1.62% within ROPE). Put differently, when controlling for factors such as *mathematics self-concept*, male students are more likely to receive grade E. Conversely, the impact of grade A on gender has a 98.89% probability of being negative, also demonstrating a significant and considerable effect (1.23% within ROPE). Moreover, all the *Rhat* values are very close to 1.000, which signifies appropriate convergence of the MCMC algorithm. These findings are supported by earlier studies, such as those conducted by Watt (2005), which indicate a tendency for male students to exhibit overconfidence in their mathematical abilities.

5.2 Significance of teaching context

In Study II, whose results are presented in Paper IV, participating students engaged with comprehensive problem situations that are both mathematically rich and contextually realistic. They practiced using problem-solving strategies inspired by George Pólya's ideas (Pólya, 1957), which bear resemblance to aspects of programming. In Paper III, this similarity is explored in a broader context than mere computer coding. Study II also illustrates an example of how programming can be integrated within the domain of mathematics education.

The key findings and contributions concerning the significance of the mathematics teaching context in understanding the interaction between general mathematics self-beliefs and student engagement are outlined below:

- *Generalized mathematics self-efficacy* significantly impacts student engagement. It serves as a mediator in two key relationships: first, between previous mathematics achievement and cognitive engagement, and second, between *mathematics self-concept* and agentic engagement.

- The relationship between an individual's self-concept in mathematics and their agentic engagement is meaningful only when accounting for factors such as previous mathematics performance, gender, and *generalized mathematics self-efficacy*.

- *Assignment anxiety* has a notable impact on the adoption of disorganized learning strategies among students. Additionally, *assignment anxiety* mediates the relationship between *generalized mathematics self-efficacy* and cognitive disengagement.

- Cognitive engagement is anticipated to increase when students utilize problem-solving strategies that bear resemblance to programming.

In the subsequent sections, I will delineate my research contributions, concentrating on the key factors influencing student engagement and their relation to the teaching context.

5.2.1 The key factor influencing engagement

The findings detailed in Paper IV reveal a significant disparity between the effects of students' *mathematics self-concept* and their *generalized mathematics self-efficacy* on their engagement with the subject. Specifically, *generalized mathematics self-efficacy* has a more pronounced impact on student engagement than *mathematics self-concept*. Consequently, this suggests that *generalized mathematics self-efficacy* also plays a significant role in shaping student motivation (Connell & Wellborn, 1991; Skinner & Belmont, 1993; Skinner et al., 2008, 2009). This observation aligns with perspectives in self-efficacy research, which suggest that self-efficacy is a more effective predictor of future motivation and performance compared to other constructs, such as self-concept (e.g., Pajares & Miller, 1994). Furthermore, the influence of prior mathematics achievement on cognitive engagement can be fully explained by its effect on *generalized mathematics self-efficacy*. This finding underscores the importance of differentiating between *generalized mathematics self-efficacy* and *mathematics self-concept*. These insights emerged from a comprehensive analysis of extensive correlations among the observed variables, as detailed in Papers I and II.

In summary, the empirical evidence highlights the importance of content-specific self-beliefs in mathematics, rather than the broader *mathematics self-concept*. Students with robust general self-efficacy beliefs in mathematics are more likely to engage more deeply in the classroom across various dimensions, including behavioral, cognitive, emotional, and agentic aspects. Furthermore, these students tend to employ disorganized learning strategies less frequently. These results support and extend previous research on the development of student motivation in educational settings (Connell & Wellborn, 1991; Skinner & Belmont, 1993; Skinner et al., 2008, 2009).

5.2.2 Factors related to the teaching context

The findings reported in Paper IV suggest that *in-class anxiety*, cognitive engagement, and agentic engagement are particularly influenced by the teaching context, especially when students are solving complex and contextually realistic mathematical problems. However, these factors appear to have only a weak correlation with one another and are influenced by different significant factors.

Regarding agentic engagement, the role of *mathematics self-concept* is particularly noteworthy, as it has been found to exert a negative influence even after accounting for factors such as *generalized mathematics self-efficacy*, assignment and *in-class anxiety*, prior mathematics achievement, and gender. This negative influence manifests through a direct effect but is counterbalanced by a positive mediating effect via *generalized mathematics self-efficacy* and *in-class anxiety*. As a result, the overall impact of *mathematics self-concept* on agentic engagement is diminished, which explains the minimal and statistically non-significant correlation coefficient observed in Paper IV.

Furthermore, the findings suggest a strong likelihood that *generalized mathematics self-efficacy* positively influences agentic engagement, both overall and when controlling for other variables. This supports the distinction between *mathematics self-concept* and *generalized mathematics self-efficacy* as empirically distinct constructs. However, the distinct predictive capabilities of these self-beliefs may have a plausible explanation: students who generally believe they are proficient in mathematics might be less motivated to express their opinions or communicate their needs, as they assume their understanding is sufficient and, therefore, engaging with the teacher is unnecessary. In contrast, course-specific competence beliefs, which are more future-oriented (e.g., "Am I able to understand the content of the mathematics course?"), may prompt students to engage more actively with their teacher to facilitate their learning.

Another noteworthy finding is that *in-class anxiety* has been identified as a unique construct, showing no significant correlations with any dimensions of engagement. Similar to *fear of failure* (see A. J. Martin & Marsh, 2003), *in-class anxiety* appears to be influenced by the teaching context, particularly the dynamics between teacher and student. Indeed, the findings in Paper IV reveal a relationship between agentic engagement and *in-class anxiety*. This is significant because agentic engagement has been associated with teachers' support for autonomy (H.-R. Jang et al., 2024). Autonomy support is a key contextual factor that promotes student engagement (Connell & Wellborn, 1991; Skinner & Belmont, 1993; Skinner et al., 2008, 2009). In other words, students who experience a caring and adaptive classroom climate are more likely to experience greater satisfaction of their autonomy, competence, and relatedness needs (Reeve, 2012; Ryan & Deci, 2020).

Regarding the interplay between cognitive engagement and the teaching context, the findings presented in Paper IV indicate a probable increase in students' cognitive engagement during Study II. However, this effect seems too minimal to hold practical significance. Despite this, the observed trend is consistent with the existing literature (e.g., Liu et al., 2023; Xing et al., 2023), a connection that might not have been identified without a comprehensive examination of the research context. Furthermore, the students' cognitive engagement was found to be strongly associated with their prior achievements in mathematics. This relationship can be attributed to their heightened levels of *generalized mathematics self-efficacy*.

General discussion

6

In this chapter, I will extend the discussion by examining the applicability of the research, its implications for teaching, and the limitations of the conducted research.

6.1 Methodology

The primary objective of this thesis is to contribute to the appropriate selection of quantitative methodologies that enhance the validity of measurements related to mathematics self-beliefs. Accordingly, I have already discussed several issues related to methodology in Chapter 3 and 4. However, in this section, I will focus on examining the broader relevance of the methodology.

This thesis presents several unique aspects. It addresses a fundamental question often overlooked in educational research, specifically concerning measurement and analysis. It provides a comprehensive examination of quantitative methodology, emphasizing the crucial role of research context in shaping data analysis and interpreting results. By adopting a pragmatic approach, I integrated various philosophical perspectives on measurement with advanced data analysis techniques. For instance, conventional statistical methods, aligned with CTT, were enhanced by incorporating statistical techniques from IRT and Bayesian frameworks. Additionally, I conducted an exploratory investigation into how general mathematics self-beliefs influence student engagement. This study aimed to illuminate potential implications for understanding student motivation.

The empirical studies demonstrate that both key aspects of this thesis (see sections 5.1 and 5.2) can be effectively addressed through a pragmatic approach to the philosophy of measurement, as discussed in Chapter 3. The pragmatic approach proves effective in addressing various challenges encountered in examining general self-beliefs in mathematics within educational research. The motivation and engagement model (Connell & Wellborn, 1991; Skinner & Belmont, 1993; Skinner et al., 2008, 2009) utilized in this thesis is an effective model that integrates various concepts related to students' motivational development. Furthermore, it offers complementary lenses for understanding and assessing student motivation. Specifically, these lenses focus on the conceptual models of students' self-beliefs in mathematics and their engagement, as well as contextual factors, such as the autonomy support provided by teachers. Moreover, the pragmatic approach underpins the research presented in this thesis, providing a framework for interpreting the findings within the research context. Additionally, I emphasize the case study methodology as an appropriate research design, particularly given the complex and underdetermined phenomena often encountered in educational research. For example, measuring psychological attributes, such as self-beliefs in mathematics, inherently involves a degree of subjectivity. To address this, I propose a procedure for carefully selecting data analysis methods suited to the specific data, while also addressing ethical considerations and making informed subjective decisions throughout the research process, with the aim of enhancing the study's epistemic objectivity.

In this thesis, I emphasize the significance of integrating qualitative reasoning into EFA rather than relying solely on quantitative heuristics. This emphasis is justified by the well-established principle that correlation does not equate to causation and, by extension, does not necessarily imply validity. Therefore, it is crucial to recognize that equating correlation with validity is problematic (Borsboom et al., 2004). Relying exclusively on a quantitative threshold and assuming that a strong correlation signifies high validity when eliminating variables from EFA can lead to unintended consequences. For example, if the correlation between two variables is extremely high, approaching a value of 1, one might infer that these variables are measuring the same underlying construct. Therefore, it could be argued that one of these highly correlated variables is redundant and might be removed.

However, it is more appropriate to conceptualize validity in qualitative terms rather than merely quantitative ones (Borsboom et al., 2004). This approach parallels the procedure in this thesis, specifically the use of qualitative reasoning in EFA before eliminating variables. For instance, in Paper I, the correlation between the items “I believe I am the kind of person who is good at mathematics” and “I believe I can think like a mathematician” was notably high. Despite the high correlation, which suggests that these variables assess the same underlying construct, they were retained due to their qualitative distinctiveness. Variables were excluded only when a clear qualitative rationale demonstrated that they measured the same underlying construct and were likely to generate identical responses from participants. An illustrative example is provided in Paper I. Specifically, the item “I get tense when I prepare for a mathematics test” was excluded from the study due to its redundancy with the item “I get nervous when taking a mathematics test”. Both items assessed the same underlying construct (test anxiety) and were considered qualitatively equivalent. Consequently, this item was removed, as respondents would likely have difficulty distinguishing between them.

The findings of this thesis are based on data collected from a upper-secondary school in Sweden, characterized by typical student achievement levels in mathematics. While the findings are specific to the context of the empirical studies, they have broader relevance and potential applicability in similar settings. The methodological insights presented are broadly applicable, though the extent of multicollinearity may vary based on the characteristics of the data, which in turn influences the complexity involved in developing a reliable factor model. Nevertheless, the four factors identified in Paper II, which pertain to students’ general self-beliefs in mathematics, are likely transferable to comparable contexts. This is mainly because each factor has been distilled into essential components of its respective construct, represented by three items. This approach may offer a more robust alternative to longer scales (Gogol et al., 2014).

However, despite respondents’ efforts to provide honest and insightful self-reports, these reports often contain inaccuracies. Paulhus and Vazire (2007) identifies several common sources of such inaccuracies. For instance, the pressure to conform to social expectations often leads individuals to provide responses they believe are socially acceptable. Additionally, some respondents may exhibit a tendency to agree with statements regardless of their true feelings, a phenomenon known as acquiescent responding. Another common inaccuracy is the tendency for some respondents to select extreme response options. Individuals from similar cultural backgrounds often display comparable self-reporting styles, which can markedly differ from those observed in other cultures. Despite the availability of several techniques designed to mitigate the impact of discrepancies during data analysis, particularly those implemented during the test construction phase (Paulhus & Vazire, 2007), it remains impossible to completely eliminate these discrepancies. Furthermore, individuals may sometimes be completely unaware of their own emotions, which can further impact the outcomes. These potential discrepancies underscore the necessity of adopting a pragmatic approach to measurement, which is employed in this thesis. From a pragmatic perspective, these inaccuracies can be viewed as integral components of a systematically organized model that serves as a lens through which we examine specific aspects of the environment (see Mislevy, 2018).

A significant limitation of the empirical studies is the sample size, which can influence the degree of multicollinearity present in the data. While increasing the sample size may help mitigate this issue, it is often impractical in educational research. Moreover, multicollinearity issues can arise even in large-scale studies. To address severe multicollinearity, one potential solution is to eliminate one or both of the variables contributing to the problem. However, this approach carries the risk of removing too many variables, which could lead to the loss of valuable information. Therefore, the findings of this thesis suggest that incorporating qualitative reasoning before removing variables may mitigate these risks and provide a more balanced solution.

6.2 Implications for mathematics teaching

The interpretation of research outcomes is intrinsically linked to the methodologies employed, ultimately providing insights into mathematics education. Specifically, this thesis utilizes a methodology that enables a detailed analysis of closely related constructs. By investigating the correlation between general mathematics self-beliefs and student engagement, this thesis offers significant insights that can enhance and inform educational practices. The empirical studies identified the levels at which students' beliefs most strongly influence their engagement. Notably, it was found that students' course-specific beliefs about mathematics, referred to as generalized mathematics self-efficacy, exert a stronger influence on the quality of their engagement than their broader mathematics self-concept. Additionally, the anxiety experienced by students regarding providing incorrect answers during class emerged as a significant factor that both distinguishes between male and female students and reduces their agentic engagement. Consequently, this factor also contributed to a reduction in their perceived support for autonomy from teachers. Given these findings, several potential implications for mathematics instruction arise.

Firstly, educators should prioritize helping students understand that arriving at the correct answer is not the sole objective of learning mathematics. Consequently, an approach to teaching that focuses exclusively on repetitive practice to achieve correct answers may be counterproductive, as it can reinforce the misconception that mathematics is solely about obtaining the correct answer, which may promote in-class anxiety. Instead, teaching methods that balance knowledge construction, which promotes conceptual understanding, with repetitive practice are more appropriate (Rittle-Johnson et al., 2001). In practice, it is essential for educators to design classroom activities that prioritize the systematic deconstruction of concepts. Additionally, teachers should allocate time to guide students through each step of the problem-solving process. This approach involves encouraging students to articulate their thought processes, clarifying the reasons behind incorrect answers, and guiding them toward correct solutions. Ultimately, this approach also influences the design of exams and assignments by emphasizing the importance of students' ability to clearly articulate their reasoning.

To reduce classroom anxiety and enhance students' autonomy, teachers should highlight the relevance of mathematics to real-world applications. This approach fosters a greater sense of agency among students (Tsai et al., 2008). Additionally, to bolster students' general self-efficacy in mathematics, educators should design activities that promote collaborative work. When students engage in joint activities, they observe their peers experiencing similar successes and challenges, which can strengthen their own course-specific beliefs in their abilities through vicarious experience. This, in turn, increases their engagement (Skinner & Belmont, 1993; Skinner et al., 2008, 2009). Recognizing that they are not alone in their struggles may also boost their motivation to actively engage with the material and seek clarification through questions, thereby advancing their learning (Skinner & Belmont, 1993; Skinner et al., 2008, 2009) and potentially reducing their in-class anxiety. In conclusion, the consistent recognition and encouragement of student efforts should be integral to the classroom learning process. Specifically, teachers should strive to develop autonomy-supportive teaching practices that involve adapting to student preferences and opinions while fostering a caring and respectful classroom environment (Reeve & Shin, 2020).

6.3 Ideas for future research

The empirical studies presented in this thesis are exploratory in nature and call for further investigation, particularly in diverse classroom settings. For example, findings suggest that in-class anxiety affects students' perceptions of autonomy support, especially in mathematics classrooms. Hypothetically, the prevalence of in-class anxiety may vary significantly across different classrooms. The results of this thesis indicate that when students experience low levels of anxiety during class, gender differences in mathematics self-belief tend to be less pronounced. Additionally, teaching practices that support student autonomy appear to be more frequently perceived in such environments. However, this hypothesis requires further investigation.

The finding that male students often exhibit overconfidence in their mathematical abilities underscores the need for further investigation into the mathematics self-concept of both male and female students. This research is particularly relevant in light of the results presented in Paper IV, which suggest that the mathematics self-concept may negatively influence students' agentic engagement. These results account for factors such as prior mathematics achievement, gender, and other related self-belief variables.

Furthermore, future research should focus on investigating the potential similarities between *fear of failure* and in-class anxiety. Although Paper IV indicates a significant relationship between agentic engagement and in-class anxiety, further empirical studies are needed to discern the differences between *fear of failure* and in-class anxiety. In particular, it would be beneficial to examine how these factors interact with mastery and performance goals, as well as to assess the impact of gender stereotypes. Moreover, the strong association between students' concerns about completing assignments at home and their reliance on disorganized learning strategies warrants further exploration. This is particularly important given that assignment anxiety has been found to mediate the relationship between generalized mathematics self-efficacy and the use of disorganized learning strategies (cognitive disengagement).

6.4 Final words

In conclusion, this research has aimed to enhance the credibility of quantitative studies by employing methodologies that improve their validity. The findings underscore the importance of in-depth data analysis within educational research, advocating for a shift away from an exclusive focus on large sample sizes, which are often difficult to achieve in this field. By drawing parallels with case study methodology, this work emphasizes the need to consider each sample as a distinct case, thereby highlighting the significance of the research context. In addition, the theoretical model for motivational development employed in this thesis is based on SDT, which emphasizes the critical role of fostering intrinsic motivation within educational contexts. This model advocates for the creation of autonomy-supportive teaching environments that enhance self-motivation, rather than relying on external pressures or rewards to stimulate engagement. Ultimately, I hope this research will inspire scholars to adopt the perspective on educational research presented herein, particularly in relation to the pragmatic interpretation of quantitative analyses. This approach has the potential to improve both the quality and the reliability of research outcomes.

References

Alfredsson, L., Björk, J., Björk, L.-E., Brolin, H., Bråting, K., Erixson, P., Heikne, H., & Palbom, A. (2019). Omfångsrika problemsituationer. In *Matematik 5000 kurs 5 blå lärobok* (pp. 242–245). Natur & kultur. (Cit. on pp. 49, 50).

Alves, M., Rodrigues, C. S., Rocha, A. M. A. C., & Coutinho, C. (2016). Self-efficacy, mathematics' anxiety and perceived importance: An empirical study with portuguese engineering students. *European Journal of Engineering Education*, *41*(1), 105–121 (cit. on pp. 18, 26).

Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools*, *45*(5), 369–386 (cit. on pp. 3, 20).

Arens, A. K., Jansen, M., Preckel, F., Schmidt, I., & Brunner, M. (2021). The structure of academic self-concept: A methodological review and empirical illustration of central models. *Review of Educational Research*, *91*(1), 34–72 (cit. on p. 14).

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science*, *11*(5), 181–185 (cit. on p. 17).

Australian Government Department of Education, Skills. (n.d.). *Digital technologies hub*. (Cit. on p. 49).

Bandura, A. (2005). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Adolescence and education* (Vol. Self-efficacy beliefs of adolescents). Greenwich, Connecticut: IAP-Information Age Publishing Inc. (Cit. on p. 18).

Bandura, A. (2008). Reconstrual of "free will" from the agentic perspective of social cognitive theory. In J. Baer, J. C. Kaufman, & R. F. Baumeister (Eds.), *Are we free?: Psychology and free will* (pp. 86–127). Oxford University Press. (Cit. on p. 13).

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*(2), 122–147 (cit. on p. 16).

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215 (cit. on pp. 15, 16).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48 (cit. on p. 56).

Baumeister, R. F. (2016). Toward a general theory of motivation: Problems, challenges, opportunities, and the big picture. *Motivation and Emotion*, *40*(1), 1–10 (cit. on p. 9).

Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research and Evaluation*, *18*(6), 1–13 (cit. on pp. 18, 52).

Berry, W. D. (1993). A formal presentation of the regression assumptions. In *Understanding regression assumptions*. SAGE Publications, Inc. (Cit. on p. 56).

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, *15*(1), 1–40 (cit. on pp. 4, 6, 15, 17–19, 59, 60).

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071 (cit. on pp. 37, 72).

Brown, T. A. (2015). *Methodology in the social sciences. confirmatory factor analysis for applied research (2nd ed.)* Guilford Press. (Cit. on p. 51).

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411 (cit. on pp. 56, 64).

Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software*, *100*(5), 1–54 (cit. on pp. 56, 64).

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28 (cit. on pp. 56, 64).

Chalmers, A. (2013). *What is this thing called science?* (4th ed.). Open University Press. (Cit. on p. 31).

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29 (cit. on p. 55).

Chen, P. Y., & Popovich, P. M. (2002). Correlation: Parametric and nonparametric measures. In M. S. Lewis-Beck (Ed.), *Quantitative applications in the social sciences* (Vol. 139). SAGE Publications. (Cit. on pp. 5, 36, 53, 54).

Choi, J., Peters, M., & Mueller, R. O. (2010). Correlational analysis of ordinal data: From pearson's r to bayesian polychoric correlation. *Asia Pacific Education Review*, *11*(4), 459–466 (cit. on p. 54).

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, *49*(12), 997–1003 (cit. on p. 43).

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge. (Cit. on pp. 5, 33, 35, 52, 59).

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes. In M. R. Gunnar & L. A. Sroufe (Eds.), *Self processes and development* (1st ed., pp. 43–77, Vol. 23). Lawrence Erlbaum Associates, Inc. (Cit. on pp. 3, 20, 21, 23, 25, 67–69, 72).

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation*, *10*(7), 1–9 (cit. on pp. 18, 52).

Deci, E. L., & Ryan, R. M. (2016). Optimizing students' motivation in the era of testing and pressure: A self-determination theory perspective. In W. C. Liu, J. C. K. Wang, & R. M. Ryan (Eds.), *Building autonomous learners: Perspectives from research and practice using self-determination theory* (pp. 9–29). Springer Singapore. (Cit. on p. 22).

Denscombe, M. (2010). *The good research guide: For small-scale social research projects* (4th ed.). Open University Press. (Cit. on p. 45).

Devine, A., Fawcett, K., Szücs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and Brain Functions*, *8*(33) (cit. on pp. 18, 26).

Ding, Y., Yang Hansen, K., & Klapp, A. (2023). Testing measurement invariance of mathematics self-concept and self-efficacy in pisa using mgcfa and the alignment method. *European Journal of Psychology of Education*, *38*(2), 709–732 (cit. on pp. 5, 61).

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*(10), 1040–1048 (cit. on pp. 10, 24).

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*(2), 256–273 (cit. on p. 13).

Eccles, J., Adler, T., & Meece, J. L. (1984). Sex differences in achievement: A test of alternate theories. *Journal of Personality and Social Psychology*, *46*(1), 26–43 (cit. on p. 27).

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 (cit. on pp. 10, 24).

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin*, *21*(3), 215–225 (cit. on p. 19).

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*(1), 109–132 (cit. on p. 19).

Elliot, A. J., McGregor, H. A., & Gable, S. (1999). Achievement goals, study strategies, and exam performance: A mediational analysis. *Journal of Educational Psychology*, *91*(3), 549–563 (cit. on p. 21).

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*(1), 103–127 (cit. on p. 26).

Field, A., Miles, J., & Field, Z. (2013). *Discovering statistics using R*. SAGE Publications. (Cit. on pp. 5, 6, 18, 52, 53, 55).

Fine, N. J. (1947). The jeep problem. *The American Mathematical Monthly*, *54*(1), 24–31 (cit. on p. 50).

Fox, J. (2022). *Polycor: Polychoric and polyserial correlations* [R package version 0.8-1]. (Cit. on p. 54).

Fredricks, J. A., Blumenfeld, P., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109 (cit. on p. 20).

Furrer, C., & Skinner, E. A. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, *95*(1), 148–162 (cit. on pp. 25, 27).

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do girls really experience more anxiety in mathematics? *Psychological Science*, *24*(10), 2079–2087 (cit. on pp. 27, 63).

Goetz, T., Cronjaeger, H., Frenzel, A. C., Lüdtke, O., & Hall, N. C. (2010). Academic self-concept and emotion relations: Domain specificity and age effects. *Contemporary Educational Psychology*, *35*(1), 44–58 (cit. on p. 17).

Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., Fischbach, A., & Preckel, F. (2014). "my questionnaire is too long!" the assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology*, *39*(3), 188–205 (cit. on p. 73).

Greene, B. A. (2015). Measuring cognitive engagement with self-report scales: Reflections from over 20 years of research. *Educational Psychologist*, *50*(1), 14–30 (cit. on pp. 21, 35).

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J., & Yeung, A. S. (2015). Expectancy-value in mathematics, gender and socioeconomic background as predictors of achievement and aspirations: A multi-cohort study. *Learning and Individual Differences*, *37*, 161–168 (cit. on p. 26).

Guyon, H., Kop, J.-L., Juhel, J., & Falissard, B. (2018). Measurement, ontology, and epistemology: Psychology needs pragmatism-realism. *Theory & Psychology*, *28*(2), 149–171 (cit. on pp. 32, 33).

Hammersley, M., & Traianou, A. (2012). *Ethics in qualitative research: Controversies and contexts*. SAGE Publications Ltd. (Cit. on pp. 32, 41, 42, 64).

Hofer, S. I., Reinhold, F., Hulaj, D., Koch, M., & Heine, J.-H. (2022). What matters for boys does not necessarily matter for girls: Gender-specific relations between perceived self-determination, engagement, and performance in school mathematics. *Education Sciences*, *12*(11) (cit. on p. 27).

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology*, *65*(1), 373–398 (cit. on p. 26).

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*(6), 581–592 (cit. on p. 26).

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*(2), 139–155 (cit. on p. 26).

ISTE. (2011). *Operational Definition of Computational Thinking: for K-12 Education* [Retrieved January 23, 2022]. (Cit. on p. 49).

Jamil, T., Ly, A., Morey, R. D., Love, J., Marsman, M., & Wagenmakers, E.-J. (2017). Default "gunel and dickey" bayes factors for contingency tables. *Behavior Research Methods*, *49*(2), 638–652 (cit. on p. 47).

Jang, H.-R., Basarkod, G., Reeve, J., Marsh, H. W., Cheon, S. H., & Guo, J. (2024). Longitudinal reciprocal effects of agentic engagement and autonomy support: Between- and within-person perspectives. *Journal of Educational Psychology*, *116*(1), 20–35 (cit. on pp. 4, 22, 25, 69).

Jang, H., Kim, E. J., & Reeve, J. (2016). Why students become more engaged or more disengaged during the semester: A self-determination theory dual-process model. *Learning and Instruction*, *43*, 27–38 (cit. on pp. 3, 20, 21, 25, 38, 40).

Jimerson, S. R., Campos, E., & Greif, J. L. (2003). Toward an understanding of definitions and measures of school engagement and related terms. *The California School Psychologist*, *8*(1), 7–27 (cit. on p. 40).

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26 (cit. on p. 29).

Kampen, J. K., & Swyngedouw, M. (2000). The ordinal controversy revisited. *Quality & Quantity*, *34*(1), 87–102 (cit. on pp. 36, 37, 54).

Kampen, J. K., & Weeren, A. (2017). A recommendation for applied researchers to substantiate the claim that ordinal variables are the product of underlying bivariate normal distributions. *Quality & Quantity*, *51*(5), 2163–2170 (cit. on p. 54).

Kerlinger, F. N., & Lee, H. B. (1999). *Foundations of behavioral research* (4th ed.). Wadsworth Publishing. (Cit. on pp. 34, 36).

Klee, H. L., Miller, A. D., & Buehl, M. M. (2022). Mathematics anxiety, self-concept, and self-efficacy: A multidimensional scaling consideration of measures. *The Journal of Experimental Education*, 1–23 (cit. on pp. 19, 59).

Kline, R. B. (2020). *Becoming a behavioral science researcher: A guide to producing research that matters* (2nd ed.). Guilford Press. (Cit. on pp. 29, 42, 43, 55, 64).

Knuth, D. E. (1984). Literate Programming. *The Computer Journal*, *27*(2), 97–111 (cit. on p. 50).

Koller, M. (2016). robustlmm: An R package for robust estimation of linear mixed-effects models. *Journal of Statistical Software*, *75*(6), 1–24 (cit. on p. 56).

Kruschke, J. K. (2018). Rejecting or accepting parameter values in bayesian estimation. *Advances in Methods and Practices in Psychological Science*, *1*(2), 270–280 (cit. on p. 66).

Lakatos, I. (1973). Science and pseudoscience [Transcript of a BBC Radio Talk broadcasted on 30 June 1973.]. (Cit. on p. 30).

Lee, H. R., McPartlan, P., Umarji, O., Li, Q., & Eccles, J. S. (2020). Just a methodological cautionary note: The jingle jangle of self-related beliefs in motivational measures. *Journal of Educational and Psychological Research*, *72*(2), 72–95 (cit. on pp. 4, 18, 59).

Lietaert, S., Roorda, D., Laevers, F., Verschueren, K., & De Fraine, B. (2015). The gender gap in student engagement: The role of teachers autonomy support, structure, and involvement. *British Journal of Educational Psychology*, *85*(4), 498–518 (cit. on p. 27).

Linnenbrink, E. A., & Pintrich, P. R. (2003). The role of self-efficacy beliefs in student engagement and learning in the classroom. *Reading & Writing Quarterly*, *19*(2), 119–137 (cit. on pp. 12, 20).

Liu, S., Peng, C., & Srivastava, G. (2023). What influences computational thinking? a theoretical and empirical study based on the influence of learning engagement on computational thinking in higher education. *Computer Applications in Engineering Education*, *31*(6), 1690–1704 (cit. on pp. 23, 34, 48, 69).

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the bayesian framework. *Frontiers in Psychology*, *10* (cit. on p. 65).

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). Bayestestr: Describing effects and their uncertainty, existence and significance within the bayesian framework. *Journal of Open Source Software*, *4*(40), 1541 (cit. on p. 65).

Mari, L., Wilson, M., & Maul, A. (2023a). Fundamental concepts in measurement. In *Measurement across the sciences: Developing a shared concept system for measurement* (2nd ed., pp. 19–48). Springer International Publishing. (Cit. on p. 36).

Mari, L., Wilson, M., & Maul, A. (2023b). Philosophical perspectives on measurement. In *Measurement across the sciences: Developing a shared concept system for measurement* (2nd ed., pp. 81–121). Springer International Publishing. (Cit. on pp. 31–33).

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, *35*(2), 63–78 (cit. on p. 3).

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, *79*(3), 280–295 (cit. on p. 15).

Marsh, H. W. (1990a). Causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology*, *82*(4), 646–656 (cit. on p. 15).

Marsh, H. W. (1990b). The structure of academic self-concept: The marsh/shavelson model. *Journal of Educational Psychology*, *82*(4), 623–636 (cit. on p. 14).

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, *23*(1), 129–149 (cit. on p. 14).

Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology*, *80*(3), 366–380 (cit. on p. 14).

Marsh, H. W., Dowson, M., Pietsch, J., & Walker, R. (2004). Why multicollinearity matters: A reexamination of relations between self-efficacy, self-concept, and achievement. *Journal of Educational Psychology*, *96*(3), 518–522 (cit. on pp. 5, 6).

Marsh, H. W., & Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, *96*(1), 56–67 (cit. on p. 14).

Marsh, H. W., & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, *47*(1), 213–231 (cit. on p. 15).

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology*, *111*(2), 331–353 (cit. on pp. 4, 6, 16–19, 59–61).

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, *20*(3), 107–123 (cit. on pp. 14, 15).

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 (cit. on p. 18).

Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology*, *89*(1), 41–54 (cit. on p. 15).

Martin, A. J., & Marsh, H. W. (2003). Fear of failure: Friend or foe? *Australian Psychologist*, *38*(1), 31–38 (cit. on pp. 61, 69).

Martin, D. P., & Rimm-Kaufman, S. E. (2015). Do student self-efficacy and teacher-student interaction quality contribute to emotional and social engagement in fifth grade math? *Journal of School Psychology*, *53*(5), 359–373 (cit. on p. 20).

Maul, A. (2013). On the ontology of psychological attributes. *Theory & Psychology*, *23*(6), 752–769 (cit. on pp. 30–32, 36).

Maxwell, J. A., & Mittapalli, K. (2010). Realism as a stance for mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social & behavioral research* (2nd ed., pp. 145–168). SAGE Publications, Inc. (Cit. on p. 30).

May, D. K. (2009). *Mathematics self-efficacy and anxiety questionnaire* [Doctoral dissertation]. University of Georgia. (Cit. on pp. 37, 39, 60, 62, 63).

McNeish, D. (2016). On using bayesian methods to address small sample problems. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(5), 750–773 (cit. on pp. 54–56, 63).

Michell, J. (2012). "The constantly recurring argument": Inferring quantity from order. *Theory & Psychology*, *22*(3), 255–271 (cit. on pp. 5, 35).

Mioevi, M. (2020). Introduction to bayesian statistics. In R. van de Schoot & M. Mioevi (Eds.), *Small sample size solutions: A guide for applied researchers and practitioners* (1st ed., pp. 3–12). Routledge. (Cit. on pp. 54, 55).

Mislevy, R. J. (2018). *Sociocognitive foundations of educational measurement* (1st ed.). Routledge: New York; London. (Cit. on pp. 32, 74).

Morán-Soto, G., & González-Peña, O. I. (2022). Mathematics anxiety and self-efficacy of mexican engineering students: Is there gender gap? *Education Sciences*, *12*(6) (cit. on pp. 18, 26).

Morey, R. D., & Rouder, J. N. (2022). *Bayesfactor: Computation of bayes factors for common designs* [R package version 0.9.12-4.4]. (Cit. on p. 47).

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, *15*, 625–632 (cit. on pp. 6, 36, 53).

Pajares, F. (2005). Gender differences in mathematics self-efficacy beliefs. In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender differences in mathematics: An integrative psychological approach* (pp. 294–315). Cambridge University Press. (Cit. on p. 26).

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*(4), 543–578 (cit. on pp. 12, 16, 18, 60).

Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology*, *24*(2), 124–139 (cit. on pp. 18, 26).

Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology*, *20*(4), 426–443 (cit. on p. 26).

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology*, *86*(2), 193–203 (cit. on pp. 6, 17, 26, 67).

Patrick, H., Ryan, A. M., & Kaplan, A. (2007). Early adolescents' perceptions of the classroom social environment, motivational beliefs, and engagement. *Journal of Educational Psychology*, *99*(1), 83–98 (cit. on p. 20).

Patrick, H., Ryan, A. M., & Pintrich, P. R. (1999). The differential impact of extrinsic and mastery goal orientations on males' and females' self-regulated learning. *Learning and Individual Differences*, *11*(2), 153–171 (cit. on p. 27).

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. Robins, R. Fraley, & R. Krueger (Eds.), *Handbook of research methods in personality psychology*. New York: Guilford. (Cit. on pp. 35, 61, 73, 74).

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*(4), 315–341 (cit. on p. 17).

Pekrun, R. (2024). Overcoming fragmentation in motivation science: Why, when, and how should we integrate theories? *Educational Psychology Review*, *36*(1), 27 (cit. on p. 19).

Pelletier, L. G., & Rocchi, M. (2016). Teachers' motivation in the classroom. In W. C. Liu, J. C. K. Wang, & R. M. Ryan (Eds.), *Building autonomous learners: Perspectives from research and practice using self-determination theory* (pp. 107–127). Springer Singapore. (Cit. on p. 22).

Peshkin, A. (1988). In search of subjectivity - one's own. *Educational Researcher*, *17*(7), 17–21 (cit. on p. 31).

Polanyi, M. (1958). Personal knowledge: Towards a post-critical philosophy. University of Chicago Press. (Cit. on p. 31).

Pólya, G. (1957). *How to solve it: A new aspect of mathematical method*. Princeton University Press. (Cit. on p. 66).

Popper, K. (1963). Conjectures and refutations: The growth of scientific knowledge. Routledge: London; New York. (Cit. on pp. 29, 30).

Putnam, H. (1974). The 'Corroboration' of Theories. In P. A. Schilpp (Ed.), *The Philosophy of Karl Popper* (pp. 221–240, Vol. 14). La Salle, Illinois : Open Court Publishing Company. (Cit. on p. 29).

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. (Cit. on p. 50).

Reeve, J. (2013). How students create motivationally supportive learning environments for themselves: The concept of agentic engagement. *Journal of Educational Psychology*, *105*(3), 579–595 (cit. on pp. 21, 22, 25, 38).

Reeve, J. (2012). A self-determination theory perspective on student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.). Springer US. (Cit. on pp. 4, 10, 12, 21, 23, 25, 69).

Reeve, J. (2018). *Understanding motivation and emotion*. John Wiley & Sons, Limited. (Cit. on pp. 3, 9, 11, 17, 21, 30, 38).

Reeve, J., Jang, H.-R., Shin, S. H., Ahn, J. S., Matos, L., & Gargurevich, R. (2022). When students show some initiative: Two experiments on the benefits of greater agentic engagement. *Learning and Instruction*, *80*, 101564 (cit. on pp. 22, 25).

Reeve, J., Jang, H., Carrell, D., Jeon, S., & Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion*, *28*(2), 147–169 (cit. on pp. 4, 22).

Reeve, J., & Shin, S. H. (2020). How teachers can support students' agentic engagement. *Theory Into Practice*, *59*(2), 150–161 (cit. on pp. 3, 4, 22, 25, 76).

Reyes, L. H. (1984). Affective variables and mathematics education. *The Elementary School Journal*, *84*(5), 558–581 (cit. on p. 14).

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? a comparison of robust continuous and categorical sem estimation methods under suboptimal conditions. *American Psychological Association*, *17*(3), 354–373 (cit. on p. 36).

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, *93*(2), 346–362 (cit. on p. 75).

Robitzsch, A. (2020). Why ordinal variables can (almost) always be treated as continuous variables: Clarifying assumptions of robust continuous and ordinal factor analysis estimation methods. *Frontiers in Education*, *5*, 1–7 (cit. on p. 53).

Rockwell, R. C. (1975). Assessment of multicollinearity - the haitovsky test of the determinant. *Sociological methods & Research*, *3*(3), 308–320 (cit. on pp. 6, 18, 59).

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36 (cit. on p. 54).

Rusch, T., Lowry, P. B., Mair, P., & Treiblmaier, H. (2017). Breaking free from the limitations of classical test theory: Developing and measuring information systems scales using item response theory. *Information & Management*, *54*(2), 189–203 (cit. on pp. 36, 37, 53, 54).

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61* (cit. on pp. 3, 9–12, 20, 22, 23, 25, 40, 69).

Sakiz, G., Pape, S. J., & Hoy, A. W. (2012). Does perceived teacher affective support matter for middle school students in mathematics classrooms? *Journal of School Psychology*, *50*(2), 235–255 (cit. on p. 20).

Schunk, D. H., & Pajares, F. (2005). Competence perceptions and academic functioning. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 85–104). Guilford Publications. (Cit. on p. 16).

Selby, C., & Woollard, J. (2014, December). *Refining an understanding of computational thinking* (Working Paper). University of Southampton. (Cit. on p. 49).

Shamoo, A. E., & Resnik, D. B. (2015). *Responsible conduct of research* (3rd). Oxford University Press. (Cit. on pp. 42, 64).

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, *46*(3), 407–441 (cit. on pp. 14, 15).

Skaalvik, E. M., & Skaalvik, S. (2002). Internal and external frames of reference for academic self-concept. *Educational Psychologist*, *37*(4), 233–244 (cit. on p. 15).

Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles*, *50*(3), 241–252 (cit. on p. 26).

Skinner, E. A. (2016). Engagement and disaffection as central to processes of motivational resilience and development. In K. R. Wentzel & D. B. Miele (Eds.), *Handbook of motivation at school* (2nd ed., pp. 145–168). Routledge. (Cit. on p. 23).

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, *85*(4), 571–581 (cit. on pp. 3, 20, 21, 23, 67–69, 72, 76).

Skinner, E. A., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology*, *100*(4), 765–781 (cit. on pp. 3, 20, 21, 23–25, 27, 38, 40, 67–69, 72, 76).

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*, *69*(3), 493–525 (cit. on pp. 4, 20, 22, 23, 37, 38, 67–69, 72, 76).

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*(2684), 677–680 (cit. on pp. 5, 36).

Street, K. E. S., Malmberg, L.-E., & Stylianides, G. J. (2022). Changes in students' self-efficacy when learning a new topic in mathematics: A micro-longitudinal study. *Educational Studies in Mathematics*, *111*(3), 515–541 (cit. on p. 48).

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics: Pearson new international edition*. Pearson Education Limited. (Cit. on pp. 5, 6, 18, 42, 52, 53, 58).

Tierney, N., & Cook, D. (2023). Expanding tidy data principles to facilitate missing data exploration, visualization and assessment of imputations. *Journal of Statistical Software*, *105*(7), 1–31 (cit. on p. 51).

Tossavainen, T., Rensaa, R. J., Haukkanen, P., Mattila, M., & Johansson, M. (2021). First-year engineering students' mathematics task performance and its relation to their motivational values and views about mathematics. *European Journal of Engineering Education*, *46*(4), 604–617 (cit. on p. 26).

Tsai, Y.-M., Kunter, M., Lüdtke, O., Trautwein, U., & Ryan, R. M. (2008). What makes lessons interesting? the role of situational and individual factors in three school subjects. *Journal of Educational Psychology*, *100*(2), 460–472 (cit. on pp. 23, 75).

Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research*, *78*(4), 751–796 (cit. on p. 16).

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, *45*(3), 1–67 (cit. on p. 51).

Vetenskapsrådet. (2017). *God forskningsed*. (Cit. on p. 41).

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804 (cit. on pp. 42–44).

Wang, J., Liu, R.-D., Ding, Y., Xu, L., Liu, Y., & Zhen, R. (2017). Teachers autonomy support and engagement in math: Multiple mediating roles of self-efficacy, intrinsic value, and boredom. *Frontiers in Psychology*, *8* (cit. on p. 20).

Watt, H. M. G. (2005). Explaining gendered math enrollments for NSW australian secondary school students. *New Directions for Child and Adolescent Development*, *2005*(110), 15–29 (cit. on pp. 27, 66).

Weiner, B. (2005). Motivation from an attribution perspective and the social psychology of perceived competence. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 73–84). Guilford Publications. (Cit. on p. 13).

Wilcox, R. R. (2010). *Fundamentals of modern statistical methods: Substantially improving power and accuracy* (2nd ed.). Springer Science+Business Media. (Cit. on p. 42).

Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing*. Elsevier Science. (Cit. on p. 53).

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological methods*, *12*(1), 58–79 (cit. on pp. 54, 55).

Wu, H., Guo, Y., Yang, Y., Zhao, L., & Guo, C. (2021). A meta-analysis of the longitudinal relationship between academic self-concept and academic achievement. *Educational Psychology Review*, *33*(4), 1749–1778 (cit. on p. 15).

Xie, F., Xin, Z., Chen, X., & Zhang, L. (2019). Gender difference of chinese high school students'math anxiety: The effects of self-esteem, test anxiety and general anxiety. *Sex Roles*, *81*(3), 235–244 (cit. on pp. 18, 26).

Xie, Y. (2016). *Bookdown: Authoring books and technical documents with R markdown*. Chapman; Hall/CRC. (Cit. on p. 50).

Xie, Y., Allaire, J., & Grolemund, G. (2018). *R markdown: The definitive guide*. Chapman; Hall/CRC. (Cit. on p. 50).

Xie, Y., Dervieux, C., & Riederer, E. (2020). *R markdown cookbook*. Chapman; Hall/CRC. (Cit. on p. 50).

Xing, W., Zhu, G., Arslan, O., Shim, J., & Popov, V. (2023). Using learning analytics to explore the multifaceted engagement in collaborative learning. *Journal of Computing in Higher Education*, *35*(3), 633–662 (cit. on pp. 23, 34, 48, 69).

Yamamoto, Y., & Miyazaki, Y. (2024). Small sample methods in multilevel analysis. *The Journal of Experimental Education*, *0*(0), 1–32 (cit. on p. 56).

Yin, R. K. (2017). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications. (Cit. on pp. 33, 43).

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79–94 (cit. on pp. 18, 52).

Young, A. J. (1997). I think, therefore i'm motivated: The relations among cognitive strategy use, motivational orientation and classroom perceptions over time. *Learning and Individual Differences*, *9*(3), 249–283 (cit. on p. 27).

Zander, L., Höhne, E., Harms, S., Pfost, M., & Hornsey, M. J. (2020). When grades are high but self-efficacy is low: Unpacking the confidence gap between girls and boys in mathematics. *Frontiers in Psychology*, *11* (cit. on p. 26).

Zeidner, M., & Matthews, G. (2005). Evaluation anxiety: Current theory and research. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 141–163). Guilford Publications. (Cit. on p. 17).

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, *25*, 82–91 (cit. on pp. 12, 20).

Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, *82*(1), 51–59 (cit. on p. 27).

Zondervan-Zwijnenburg, M., Peeters, M., Depaoli, S., & Van de Schoot, R. (2017). Where do priors come from? applying guidelines to construct informative priors in small sample research. *Research in Human Development*, *14*(4), 305–320 (cit. on p. 56).

List of Figures

List of Tables

Summary in Swedish

Självförtroende i matematik spelar en avgörande roll för hur elever uppfattar sig själva och påverkar deras beteende i klassrummet, och deras engagemang och motivation. Motivationsforskningen lider dock av en överskott av liknande begrepp, vilket kan leda till höga korrelationer och försvåra tolkningen av resultaten. Tilltron till sin matematiska förmåga är ofta mer allmänt definierad och saknar specifika prestationskriterier, till skillnad från tilltron att klara av specifika uppgifter. För att stärka trovärdigheten i förutsägelser om matematikprestationer är det avgörande att tydligt särskilja mellan olika typer av självförtroende. Denna avhandling syftar till att förfina metoder för att mäta dessa föreställningar och därmed ge nya insikter.

En utforskande faktoranalys av data från gymnasieelever utfördes med både kvantitativa metoder och kvalitativa resonemang, vilket avslöjade en tydlig distinktion mellan självuppfattning i matematik och allmän tilltro till matematiska förmågor. En nyckelskillnad mellan dessa två konstruktioner är att den allmänna tilltron har en större påverkan på elevernas engagemang. Faktoranalysen identifierade även två konstruktioner relaterade till matematikångest, där en faktor, oron för att ge felaktiga svar i klassrummet, visade stora skillnader mellan män och kvinnor och negativt påverkade elevernas engagemang för autonomi. Sammantaget understryker dessa fynd vikten av att eliminera variabler med hög interkorrelation från analysen, förutsatt att denna procedur stöds av kvalitativa resonemang. Resultaten indikerar behovet av ett pragmatiskt tillvägagångssätt för att undersöka självförtroende i matematik och erbjuder nya insikter om den betydelsefulla rollen av autonomi-stödjande undervisning i att forma elevers matematiska självförtroende.

Appendix A

LULEÅ
TEKNISKA
UNIVERSITET

Samtycke till deltagande i studie

Hej, jag heter Erik Bergqvist och är doktorand i matematik och lärande på Luleå tekniska universitet. Jag planerar genomföra en enkätstudie på gymnasieskolan. Syftet är att undersöka några faktorer som påverkar motivationen i matematik och dess koppling till studieresultat. Deltagandet är frivilligt och du kan närsomhelst välja att avbryta din medverkan. Det är viktigt att tänka på att det inte är ett prov och kommer alltså inte bedömas, och ditt deltagande kommer inte påverka ditt betyg i matematik.

Om du väljer att delta i studien kommer du först få svara på några bakgrundsfrågor eftersom dina svar behöver kopplas ihop med ditt studieresultat i matematik. I övrigt innehåller enkäten ett antal påståenden som du kommer att få ta ställning till hur väl det stämmer in på dig. Webbadress till enkäten finns längst ner på denna sida. Jag tillsammans med mina handledare (Maria Johansson och Timo Tossavainen) kommer att vara de enda personerna som har tillgång till dina svar. All den information som kommer fram kommer bara att användas för denna studie, och kommer att behandlas med försiktighet. Resultatet kommer att publiceras i vetenskapliga artiklar och på konferenser på ett sådant sätt att ingen deltagare kan identifieras.

Vänligen kryssa för ett av följande två alternativen (JA / NEJ):

☐ JA, jag samtycker till medverkan i studien.

☐ NEJ, jag samtycker <u>inte</u> till medverkan i studien.

....	....
Underskrift	Namnförtydligande

Webbadress till enkät: https://tinyurl.com/y4t3uho6

Eller skanna QR-kod:

Erik Bergqvist, doktorand matematik och lärande
Luleå tekniska universitet, 971 87 Luleå
0920-49 10 00 www.ltu.se

Figure A.1. Informed consent

Table A.1. Items from MSEAQ with the original wording in italics

General mathematics self-efficacy
Jag känner mig tillräckligt självsäker för att ställa frågor under matematiklektionen. *I feel confident enough to ask questions in my mathematics class.*
Jag tror det ska gå bra på matematikprov. *I believe I can do well on a mathematics test.*
Jag tror jag kommer att bli klar med alla uppgifter i matematikkursen. *I believe I can complete all of the assignments in a mathematics course.*
Jag tror jag är en person som är bra på matematik. *I believe I am the kind of person who is good at mathematics.*
Jag tror det kommer att gå bra när jag behöver använda mina matematikkunskaper i mitt framtida yrke. *I believe I will be able to use mathematics in my future career when needed.*
Jag tror jag kan förstå matematikkursens innehåll. *I believe I can understand the content in a mathematics course.*
Jag tror jag kan få betyget A i matematikkursen. *I believe I can get an "A" when I am in a mathematics course.*
Jag tror det ska gå bra att lära mig matematik. *I believe I can learn well in a mathematics course.*
Jag känner mig självsäker när jag gör matematikprov. *I feel confident when taking a mathematics test.*
Jag tror jag är en person som kan använda matematik. *I believe I am the type of person who can do mathematics.*
Jag känner att det kommer att gå bra i nästa matematikkurs. *I feel that I will be able to do well in future mathematics courses.*
Jag tror jag kan använda matematiken i kursen. *I believe I can do the mathematics in a mathematics course.*
Jag tror jag kan tänka som en matematiker. *I believe I can think like a mathematician.*
Jag är självsäker när jag använder mina matematikkunskaper utanför skolan. *I feel confident when using mathematics outside of school.*
Mathematics anxiety
Jag blir spänd när jag förbereder mig inför ett matematikprov. *I get tense when I prepare for a mathematics test.*
Jag blir nervös när jag är tvungen att använda mina matematikkunskaper utanför skolan. *I get nervous when I have to use mathematics outside of school.*
Jag oroar mig för att det inte kommer att gå bra när jag behöver använda mina matematikkunskaper i mitt framtida yrke. *I worry that I will not be able to use mathematics in my future career when needed.*
Jag oroar mig för att jag inte kommer att få ett bra betyg i matematikkursen. *I worry that I will not be able to get a good grade in my mathematics course.*
Jag oroar mig för att det inte kommer att gå bra på matematikproven. *I worry that I will not be able to do well on mathematics tests.*
Jag känner mig stressad av att lyssna på min matematiklärare. *I feel stressed when listening to mathematics instructors in class.*
Jag blir nervös när det ställs frågor under matematiklektionen. *I get nervous when asking questions in class.*
Jag blir stressad av att jobba med matematikuppgifter hemma. *Working on mathematics homework is stressful for me.*
Jag oroar mig för att jag inte kommer att kunna tillräckligt mycket för att göra bra ifrån mig i nästa matematikkurs. *I worry that I do not know enough mathematics to do well in future mathematics courses.*
Jag oroar mig för att jag inte kommer att bli klar med alla uppgifter i matematikkursen. *I worry that I will not be able to complete every assignment in a mathematics course.*
Jag oroar mig för att jag inte kommer att förstå matematiken som finns med i kursen. *I worry I will not be able to understand the mathematics.*
Jag oroar mig för att jag inte kommer att få betyget A i matematikkursen. *I worry that I will not be able to get an "A" in my mathematics course.*
Jag oroar mig för att jag inte kommer att lära mig bra i matematikkursen. *I worry that I will not be able to learn well in my mathematics course.*
Jag blir nervös när jag gör matematikprov. *I get nervous when taking a mathematics test.*
Jag är rädd för att säga fel under matematiklektionen. *I am afraid to give an incorrect answer during my mathematics class.*

Table A.2. Engagement and disengagement items with the original wording in italics

Behavioral engagement/disengagement
Jag lyssnar ordentligt. *When I'm in this class, I listen very carefully.*
Jag är uppmärksam. *I pay attention in this class.*
Jag försöker göra mitt bästa för att det ska gå bra för mig. *I try hard to do well in this class.*
Jag jobbar så hårt jag bara kan. *In this class, I work as hard as I can.*
Jag deltar i klassrumsdiskussioner. *When I'am in this class, I participate in class discussions.*
Jag låtsas att jag jobbar under lektionerna. *When I'm in this class, I just act like I'm working.*
Jag anstränger mig inte speciellt mycket. *I don't try very hard in this class.*
Jag gör precis det som krävs för att klara mig. *In class, I do just enough to get by.*
Jag tänker på helt andra saker. *When I'm in class, I think about other things.*
Cognitive engagement
Jag försöker beskriva matematiska begrepp med egna ord. *When reading for this class, I try to explain the key concepts in my own words.*
Jag försöker sammanfatta nya moment i kursen med egna ord. *When learning about a new topic in this course, I usually try to summarize it in my own words.*
Jag försöker koppla ihop det jag lär mig med det jag redan kan. *When reading for this class, I try to connect the ideas I am reading about with what I already know.*
Jag försöker skapa egna exempel som ska hjälpa mig att förstå matematiska begrepp. *When thinking about the concepts in this class, I try to generate examples to help me understand them better.*
Cognitive disengagement
Jag vet inte vad jag ska plugga eller var jag ska börja. *In this course, I often find that I don't know what to study or where to start.*
Jag är osäker på hur jag ska plugga. *I'm not sure how to study for this course.*
Jag tycker det är svårt att plugga på ett tidseffektivt sätt. *In this course, I find it difficult to organize my study time effectively.*
Jag har svårt att komma på hur jag ska göra för att lära mig matematik. *When I study for this course, I have trouble figuring out what to do to learn the material.*
Agentic engagement
Jag ser till att min lärare vet vad jag vill och behöver. *I let my teacher know what I need and want.*
Jag uttrycker mina åsikter och vad jag helst vill göra. *During this class, I express my preferences and opinions.*
Jag frågar läraren när jag behöver något. *When I need something in this class, I'll ask the teacher for it.*
Jag ställer frågor för att lära mig bättre. *During class, I ask questions to help me learn.*
Jag ser till att min lärare vet vad jag tycker är intressant. *I let my teacher know what I am interested in.*
Emotional engagement/disengagement
Jag är intresserad av det vi jobbar med på lektionen. *When we work on something in this class, I feel interested.*
Jag tycker det är roligt på lektionen. *This class is fun.*
Jag gillar att lära mig nya saker under lektionen. *I enjoy learning new things in this class.*
Jag mår bra på lektionen. *When I'm in this class, I feel good.*
Jag är engagerad i det vi jobbar med på lektionen. *When we work on something in this class, I get involved.*
Jag känner mig uttråkad när vi jobbar med något på lektionen. *When we work on something in class, I feel bored.*
Jag mår dåligt på lektionen. *When I am in this class, I feel bad.*
Jag känner mig orolig på lektionen. *When I'm in this class, I feel worried.*
Jag tycker inte att det är roligt på lektionen. *This class is no fun for me.*
Jag känner mig uppgiven och har svårt att jobba på lektionen. *When we work on something in class, I feel discouraged.*

Table A.3. The 17 Items from MSEAQ analyzed in Paper II, with the original wording in italics

General mathematics self-efficacy
Jag känner mig tillräckligt självsäker för att ställa frågor under matematiklektionen. *I feel confident enough to ask questions in my mathematics class.*
Jag tror det ska gå bra på matematikprov. *I believe I can do well on a mathematics test.*
Jag tror jag kommer att bli klar med alla uppgifter i matematikkursen. *I believe I can complete all of the assignments in a mathematics course.*
Jag tror jag är en person som är bra på matematik. *I believe I am the kind of person who is good at mathematics.*
Jag tror det kommer att gå bra när jag behöver använda mina matematikkunskaper i mitt framtida yrke. *I believe I will be able to use mathematics in my future career when needed.*
Jag tror jag kan förstå matematikkursens innehåll. *I believe I can understand the content in a mathematics course.*
Jag tror jag kan få betyget A i matematikkursen. *I believe I can get an "A" when I am in a mathematics course.*
Jag tror det ska gå bra att lära mig matematik. *I believe I can learn well in a mathematics course.*
Jag känner mig självsäker när jag gör matematikprov. *I feel confident when taking a mathematics test.*
Jag tror jag är en person som kan använda matematik. *I believe I am the type of person who can do mathematics.*
Jag tror jag kan använda matematiken i kursen. *I believe I can do the mathematics in a mathematics course.*
Jag tror jag kan tänka som en matematiker. *I believe I can think like a mathematician.*
Mathematics anxiety
Jag blir nervös när det ställs frågor under matematiklektionen. *I get nervous when asking questions in class.*
Jag blir stressad av att jobba med matematikuppgifter hemma. *Working on mathematics homework is stressful for me.*
Jag oroar mig för att jag inte kommer att bli klar med alla uppgifter i matematikkursen. *I worry that I will not be able to complete every assignment in a mathematics course.*
Jag blir nervös när jag gör matematikprov. *I get nervous when taking a mathematics test.*
Jag är rädd för att säga fel under matematiklektionen. *I am afraid to give an incorrect answer during my mathematics class.*

Table A.4. Summary of MCAR-test results for sample (#) 1, 2, and 3

#	N	Math self-efficacy and anxiety	Engagement
1	79	$\chi^2(196.00) = 190.98,\ p = 0.59$	
2	77	$\chi^2(80.00) = 99.59,\ p = 0.07$	$\chi^2(206.00) = 189.64,\ p = 0.79$
3	41	$\chi^2(64.00) = 67.13,\ p = 0.37$	$\chi^2(105.00) = 109.96,\ p = 0.35$

Table A.5. Summary of χ^2 bivariate statistics for study 1

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	22	14	14	26*	25	17	27*	14	22	9	17	14	18	25	19	21
		49	12	18	22	41	30*	11	21	17	16	9	16	28*	30*	20
			14	19	20	27*	16	37	12	15	14	13	14	26*	20	36
				13	24	18	30*	15	20	17	18	20	21	24	18	16
					19	27*	13	11	15	16	12	16	17	12	18	27*
						21	22	14	17	15	24	19	15	21	11	13
							31	23	45	22	32	16	16	12	20	31*
								17	19	12	14	20	23	15	11	25*
									19	19	20	13	12	16	15	25
										22	7	11	19	20	16	16
											14	13	16	13	17	25
												28*	21	16	26*	18
													40	16	8	30*
														31*	17	20
															16	11
																20

Note: N = 155. Significance levels: $p < 0.001$, $p < 0.01$, * $p < 0.05$ (two-tailed)

Table A.6. Summary of χ^2 bivariate statistics for engagement items (sample 2)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
	9	20	23	19	14	19	10	23*	15	10	26*	13	8	11	12	8	17	15	13	18	15	17
		13	7	13	9	15	4	3	8	9	8	8	3	9	11	16	18*	10	13	20*	15	16
			12	9	22	12	11	17	5	12	25	23	20	10	17	12	16	19	12	14	20	15
				9	15	11	17	25	21*	21	18	18	22	9	18	16	21*	15	12	21	14	9
					16	8	16	16	10	14	14	24	13	10	10	12	12	22	16	25	12	9
						16	9	20*	21*	7	16	15	9	6	12	19	26	21	15	13	19	26*
							17	18	7	19	10	25	25	19	17	19	13	24	17	18	9	23
								1	9	8	13	21	12	12	12	22	11	13	20	22	21	21
									12	10	20*	7	7	10	11	8	12	5	12	25	14	14
										16	25	14	12	16*	11	10	9	14	25*	18	11	18
											12	14	11	12	11	18	12	12	16	16	22	21
												19	14	21*	17	13	11	8	10	19	18	20
													20	12	14	22	9	21	12	19	20	25
														16	13	10	5	19	17	20	22	16
															7	8	13	17	20*	15	16	21*
																13	7	31	19	10	24	12
																	12	23	10	16	12	13
																		10	11	16	15	10
																			16	16	13	14
																				17	23	22
																					25*	16
																						26*

Note: N = 76. Significance levels: $p < 0.001$, $p < 0.01$, * $p < 0.05$ (two-tailed)

Colophon

This thesis was typeset with $\LaTeX\,2_\varepsilon$. It uses the *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the *Clean Thesis* style at http://cleanthesis.der-ric.de/.

Declaration

The work has been completed solely with the help of references. ChatGPT (version 3.5) was used for editing assistance to check expressions for clarity, grammar, spelling, and consistency in language use.

Erik Bergqvist
Luleå, September 2024

Department of Department of Health, Education and Technology
Division of Education and Languages

ISSN 1402-1544
ISBN 978-91-8048-614-9 (print)
ISBN 978-91-8048-615-6 (pdf)

Luleå University of Technology 2024

LULEÅ
UNIVERSITY
OF TECHNOLOGY