





Measuring Math Anxiety Through Self-Reports and Physiological Data

Febe Demedts^{1,2} , Jan Cornelis³, Bert Reynvoet⁴ , Delphine Sasanguie⁵ , Fien Depaepae^{1,2} 

[1] Centre for Instructional Psychology and Technology, Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium. [2] imec research group itec, KU Leuven, Leuven, Belgium. [3] imec, Leuven, Belgium. [4] Brain & Cognition, Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium. [5] Research Centre for Learning in Diversity, HOGENT, Gent, Belgium.

Journal of Numerical Cognition, 2023, Vol. 9(3), 380–397, <https://doi.org/10.5964/jnc.9735>

Received: 2022-06-20 • Accepted: 2023-09-13 • Published (VoR): 2023-11-30

Handling Editor: Andreas Obersteiner, Technical University Munich, Munich, Germany

Corresponding Author: Febe Demedts, Faculty of Psychology and Educational Sciences, KU Leuven campus Kulak Kortrijk, E. Sabbelaan 51, 8500 Kortrijk, Belgium. E-mail: febe.demedts@kuleuven.be

Supplementary Materials: Materials [see [Index of Supplementary Materials](#)]



Abstract

Math anxiety (MA) is an important affective factor that contributes to individuals' math proficiency. While self-reports are commonly used to measure MA, a number of limitations are inherently connected to this measuring method. Physiological responses are considered a promising alternative approach, but research is scarce and the empirical evidence is scattered. Therefore, this paper aimed to (1) investigate whether different types of tasks (i.e., difficulty and topic) result in differences regarding self-reported anxiety and physiological measures, and (2) analyse whether physiological measures can account for differences in self-reported MA. We manipulated the difficulty level of a math and non-math task, so this study had a two-by-two experimental within-subject design. The participants were 44 undergraduate students. In terms of the first research aim, results revealed that the difficult math task elicited more self-reported anxiety compared to the easy math task and the difficult non-math task. However, these differences are barely detected by physiological measures. Regarding the second research aim, results showed that phasic galvanic skin responses and heart coherence ratio significantly predicted the self-reported MA. Our findings point to a possible contribution of using physiological measures to understand the construct of MA, meanwhile warning for a too optimistic use of this measurement method.

Keywords

math anxiety, physiological responses, skin temperature, galvanic skin response, heart rate, heart rate variability, mathematical proficiency

Theoretical Framework

Feelings of math anxiety (MA) can have long-lasting effects on individuals' personal lives and can hamper the development of mathematical proficiency (Dowker et al., 2016). Accurately measuring MA is crucial for educational research to examine this phenomenon in depth. Research on MA claims a distinction between trait-MA and state-MA (Conlon et al., 2021; Orbach et al., 2019; Strohmaier et al., 2020). Whereas trait-MA refers to a person's tendency of being anxious when confronted with mathematics, state-MA is considered as a fluctuating feeling of anxiety in response to specific mathematical situations (Dowker et al., 2016; Harley, 2016; Pekrun & Bühner, 2014). This is exemplified by research observing that state-MA can be influenced by various task characteristics, such as the topic (Doz et al., 2023) or the difficulty level (Artemenko et al., 2015; Trezise & Reeve, 2018) of a task. Self-report questionnaires are commonly used to measure both trait- and state-MA (Cipora et al., 2019). However, in order to measure state-MA, studies have shown that self-report measures are less suited, but that new approaches such as physiological responses could offer solace.



Because, so far, very little research has been done on the use of physiological responses to measure MA (Horvers et al., 2021), this study aims to explore the use of physiological measures as indicators for students' feelings of state-MA. In the current study, the task difficulty will be manipulated to investigate differences in state-MA for an easy compared to a difficult task. This study represents an impetus in using physiological measures to acquire a more profound understanding of the concept of MA.

Math Anxiety

Over the years, multiple definitions of MA have been formulated, highlighting different aspects of this concept. A highly cited definition is the one by Richardson and Suinn (1972) who state that “math anxiety involves feelings of tension and anxiety that interfere with the manipulation of numbers and the solving of mathematical problems in a wide variety of ordinary life and academic situations” (p. 551). This definition emphasizes the domain-specificity of MA, pointing out that MA appears to be specifically related to mathematical situations; whereas studies have shown a moderate correlation between general anxiety and MA, these types of anxiety are considered to be distinct (Hill et al., 2016; Mammarella et al., 2019). No uniform prevalence rates of MA are available, as these are highly dependent on the sample and criteria used to define ‘highly math-anxious individuals’, but it is unmistakable that many people experience MA in different cultures and across the world (Dowker et al., 2016).

Many studies have focused on the relationship between MA and math achievement, indicating significant negative correlations between these constructs (e.g., Namkung et al., 2019). However, there is no consensus about the causality (Carey et al., 2016). Presumably, there is a reciprocal, reflective relation between MA and math achievement, meaning that poor math performance might generate feelings of MA, which in turn might lead to worse math performance (Carey et al., 2016; Cargnelutti et al., 2017; Ramirez et al., 2018). Nevertheless, the relation between MA and math achievement indicates the importance of this negative emotional reaction and its impact on mathematical proficiency.

As already mentioned, research on MA distinguishes between state-MA and trait-MA (Conlon et al., 2021; Orbach et al., 2019; Strohmaier et al., 2020) and so far, research mainly focused on trait-MA (Dowker et al., 2016). Only in recent years state-MA received increasing research attention, indicating that MA can vary depending on the difficulty (Trezise & Reeve, 2018) and the topic of the task (Halme et al., 2022). This can be related to the notion of ‘the anxiety-complexity effect’, which refers to the fact that increasing task demands can induce more feelings of MA (Artemenko et al., 2015; Suárez-Pellicioni et al., 2013). Moreover, recent research suggests a relationship between MA and perceived difficulty of task-specific elements, such that students who experience MA rate math-related tasks as more complex and demanding (Doz et al., 2023).

Measuring Math Anxiety Through Self-Reports

A vast majority of studies have utilized self-reports to measure MA in adolescents and adults (Cipora et al., 2019; Dowker et al., 2016). Much research has focused on developing valid and reliable self-report questionnaires to measure MA (Dowker et al., 2016). The use of these self-report questionnaires fits into the predominant research focus on trait-MA, in which MA is considered a stable trait of a person. Many valuable insights regarding MA are derived from the use of these self-reports in studies that investigated the prevalence of MA, the development of MA, and the relationship between MA and math achievement.

The practice of measuring MA is being questioned because of recent studies that have revealed a discrepancy between trait-MA and state-MA (e.g., Orbach et al., 2019, 2020), since state-MA is a situation-specific anxiety reaction which can fluctuate before, during, and after a particular math context (Cipora et al., 2019; Orbach et al., 2019). In an attempt to measure state-MA through self-reporting, the Single-Item Math Anxiety questionnaire (SIMA; Ashcraft, 2002; Núñez-Peña & Suárez-Pellicioni, 2014) was developed to quickly determine how participants experience MA in a specific math-related context by asking participants “On a scale from 1 to 10, how math anxious are you?”. While good psychometric properties were found for the SIMA, some fundamental disadvantages can be noted from a methodological perspective when self-reports are used to measure state-MA.

First, self-reports inherently require participants to reflect and evaluate their experiences retrospectively. Yet, it is conceivable that these types of answers that depend on an individual's trustworthiness and accuracy might lead to false conclusions (Dowker et al., 2016). In addition, this requirement of being able to reflect retrospectively might be met

for adults and adolescents, because they can verbally express their mental affective state (Cipora et al., 2019; Pekrun & Bühner, 2014). However, this is not viable for all target groups, since metacognitive skills might differ with age and abilities, making these self-reports less feasible for young children or people with disabilities (Mammarella et al., 2019). Second, self-reports are always in some way obtrusive for the learner and to the learning process (Pekrun & Bühner, 2014). Essentially, self-reports imply that the focus of the learner is shifted from the learning task to the self-report measure which prevents continuous questioning of state-MA during the learning process (Harley, 2016). Therefore, studies typically question state-MA at the end of a (set of) task(s) (e.g., Orbach et al., 2019; Strohmaier et al., 2020). This might be problematic, as studies indicate significantly different state-MA ratings based on the timepoint at which state-MA was measured (i.e., before, during, or after a math task) (Conlon et al., 2021).

Given those disadvantages, the use of objective measures, like for instance log traces that automatically track user activity or record physiological reactions during learning, in addition to more subjective measurements (i.e., self-reports) for capturing MA is recommended (Horvers et al., 2021; Noroozi et al., 2020). The use of multimodal measures can be an attempt to overcome limitations of a single measurement method to obtain a more in-depth understanding of the phenomenon of MA (Horvers et al., 2021).

Measuring Math Anxiety Through Physiological Data

Affective factors during learning are typically associated with physiological arousal (Braithwaite et al., 2015; Harley, 2016; Kreibig, 2010). Analysing these physiological reactions to measure MA can be considered as an attempt that meets the limitations of self-reports because these responses can be mapped in a more continuous, unobtrusive, and objective manner (Giannakakis et al., 2022; Roos et al., 2021).

Some researchers endeavoured to analyse physiological responses in mathematical situations, ranging from neuro-imaging methods to signals associated with the activation of the autonomic nervous system (Dowker et al., 2016; Suárez-Pellicioni et al., 2016). Recent research in the field of MA making use of neuro-imaging suggested that fear and pain areas in the brain are activated when confronted with mathematics (i.e., the right basolateral amygdala regarding the fear network and the bilateral dorso-posterior insula and mid-cingulate cortex regarding the pain network; Artemenko et al., 2015; Suárez-Pellicioni et al., 2016). Although brain imaging addresses some of the limitations of self-reports and can yield interesting results, this method is hard to adopt in research within an ecologically valid classroom context. Another interesting approach to identify affective states are physiological responses that are driven by the activation of the autonomic nervous system (Cipora et al., 2019; Dowker et al., 2016). Recent advances in continuous and unobtrusive sensing can be promising in this regard. For instance, wrist-worn wearables are viable for monitoring physiological responses in learning environments (Poh et al., 2010).

In particular, studies on anxiety and stress generally indicate increased galvanic skin response (GSR), increased heart rate (HR), decreased heart rate variability (HRV), and decreased skin temperature (ST) (Kreibig, 2010; Smets et al., 2018). GSR can be defined as the autonomic changes in the electrical properties of the skin (Braithwaite et al., 2015). Typically, a tonic and phasic component are dissociated. The tonic component (i.e. skin conductance level; SCL) reflects the gradually changing background activity, whereas the phasic component (i.e. skin conductance phasic; SCPh) exhibits fast reactivity (Boucsein, 2012; Braithwaite et al., 2015). A considerable amount of literature has been published on the relationship between responses of the autonomic nervous system and anxiety (Giannakakis et al., 2022; Horvers et al., 2021; Kreibig, 2010; Mertens et al., 2017; Roos et al., 2021), although studies on MA-specific reactions are limited and have yielded scattered results.

In general, studies on MA including physiological data can be divided according to their objective (Horvers et al., 2021). First, some studies examine how physiological responses vary during the learning process or across different tasks. These studies typically manipulate the difficulty level of mathematical tasks to investigate whether physiological arousal increases as a function of problem difficulty. For instance, in the study of Singh and colleagues (2019), two mental arithmetic tasks varying in difficulty level were created. The participants were undergraduate students aged between 18 and 25 years. Results showed that the mean HR was significantly lower in the difficult task, which was related to mental stress. Similarly, in the study of Hunt and colleagues (2017), where children aged between 9 and 11 years participated, the findings indicated a significant decrease in HR between a baseline and two-digit math problems. However, the decrease in HR between baseline and more difficult three-digit math problems was not significant.

The second category of studies investigates the relationship between physiological responses and self-report measures of MA. Among the earliest experimental designs was the one by Dew and colleagues (1984), who administered three sets of mathematical problems to undergraduate students. Unfortunately, no information was provided on the exact age of participants and the technology used to measure physiological responses. Two significant relationships were found between trait-MA and physiological responses, notably a significant positive correlation with the tonic GSR and a significant negative correlation with mean HR. The direction of the relationship of these results are in contrast with the findings of Qu and colleagues (2020), who examined MA in slightly younger children compared to the study of Dew and colleagues (1984), namely students between 15 and 17 years old. Qu and colleagues (2020) used a trait-MA questionnaire on the one hand and on the other hand measured GSR and HR with wrist-worn wearables before and during a math exam. They observed that the self-reported trait-MA, was negatively correlated with the tonic GSR in the 5-min period before the exam specifically. Furthermore, the self-reported trait-MA was positively correlated with the mean HR during the exam. Still, other studies failed to observe significant correlations between (mostly trait) MA measured by self-reports and physiological responses (Hunt et al., 2017; Strohmaier et al., 2020). Different from the previous studies investigating the association between trait-MA and physiological responses, Strohmaier and colleagues (2020) were the first attempting to correlate physiological measurements measured with an Empatica E4 wristband with self-reported state-MA of undergraduate students with a mean age of 23. However, again no significant correlations were observed. Although these mixed findings might be a result of differences between studies in terms of the age of participants, design of the tasks, and/or the technology through which the physiological responses are measured, we can conclude that the picture of using physiological measures is extremely scattered at this moment.

Challenges in Research on the Association Between Self-Reported and Physiological Approaches to Assess MA

To date, the relationship between self-reported MA and various physiological measures remains equivocal. Although some studies have already explored the possibility to use physiological measures in relation to MA, this has led to inconsistent results. Moreover, several research gaps can be identified concerning the design and the methodology of previous studies.

The first limitation is associated with the experimental design of previous studies. Given the evidence that the difficulty of a math task is related to MA (Artemenko et al., 2015; Trezise & Reeve, 2018), task difficulty was manipulated in research designs based on the hypothesis that more difficult math tasks would evoke more physiological arousal (Hunt et al., 2017). The studies by Singh and colleagues (2019) and Hunt and colleagues (2017) addressed this hypothesis and offered two math tasks where the difficulty level differed according to the required calculations or the number of digits, respectively. Although these studies manipulated the difficulty level of math tasks, these studies did not control for perceived difficulty by the participants as crucial factor in the relationship with MA (Doz et al., 2023). Moreover, such studies with two math tasks with different levels of difficulty or studies that compare an empty baseline condition with a math task cannot ascertain whether physiological arousal is due to the difficulty of the task or the specific mathematics context.

The second limitation is related to the association between physiological and self-reported data. Several studies correlated physiological measures (a proxy of state-MA) with self-reported trait-MA (for example, Dew et al., 1984; Qu et al., 2020), whereas the discrepancy between state-MA and trait-MA has been well established (Bieg et al., 2014; Orbach et al., 2020). Therefore, it is important to examine physiological measures in relation to state self-reports. Although Strohmaier and colleagues (2020) measured self-reported state-MA, they only included one physiological measure (i.e., phasic GSR) and thereby did not pursue an advanced multimodal approach. Combining multimodal data can provide new valuable insights in terms of their interrelationship and effects on learning (Horvers et al., 2021; Noroozi et al., 2020).

The aim of this study is therefore to tackle these research gaps by using a two-by-two experimental design manipulating both task difficulty (easy vs difficult) and topic (math vs non-math). This leads to four task conditions for which we measured self-reported state anxiety, perceived difficulty and various physiological responses (GSR, HR, HRV, and ST).

The following research questions are addressed:

RQ1. Does the difficulty level (easy vs difficult) and topic (math vs non-math) result in differences in terms of ...?

- a. self-reported state (math) anxiety
- b. physiological measures: skin temperature, heart rate (variability), and galvanic skin response

RQ2. Can physiological measures account for differences in self-reported state math anxiety?

Method

Participants

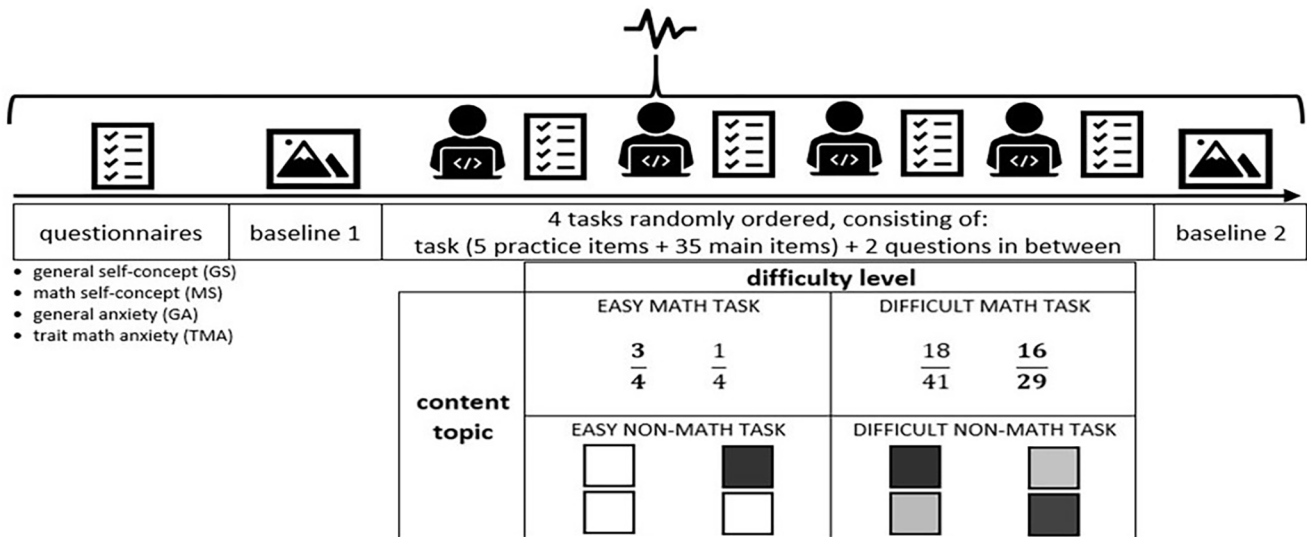
Participants were 44 undergraduate students ($M_{age} = 20.73$, $SD_{age} = 4.81$, 84% female) in ‘Educational Sciences’ within the Faculty of Psychology and Educational Science in Belgium. All students were invited by email and participation was voluntary on the basis of active written informed consent in exchange for course credits or a gift voucher. The data collection was conducted in December 2020, according to the regulations of and approved by the ethical committee of KU Leuven.

Procedure

Figure 1 illustrates the entire procedure. During the experiment, appropriate safety measures were maintained to minimize the risk of contamination by Covid-19.

Figure 1

Visualisation of the Study Design



The study was conducted in a laboratory setting to control for as many as possible confounding variables. The researcher was present while the participant went through the experiment individually, using a presentation with step-by-step guidelines. After attaching the devices for capturing the physiological responses, the student received four questionnaires through the online environment Qualtrics. All questionnaires had to be answered using a Likert scale, with higher scores indicating higher levels of self-concept or anxiety. First, the Academic Self-concept Scale (ASC scale; Liu & Wang, 2005) was used to measure general self-concept. Second, we used the Math subscale of the Self-Description Questionnaire III (SDQ-III; Marsh & O’Neill, 1984) to assess the level of math self-concept. Third, we measured general anxiety with the Generalized Anxiety Disorder (GAD7; Spitzer et al., 2006). Finally, the Abbreviated Math Anxiety Scale

(AMAS; Hopko et al., 2003) was used to measure trait-MA. These instruments had a Cronbach's alpha of respectively, .74, .88, .82, and .88. These questionnaires were administered, as this study is part of a larger research project, but not analysed since this was not the scope of this paper. However, the timeframe to complete the questionnaires (+/- six minutes) was used as a period of habituation of the devices to the participant's body.

Afterwards, a baseline measurement was conducted in which the participant was in a quiet condition and watched a video with nature images and sounds. In the main part of the experiment four tasks had to be performed, which were offered in random order. After completing the tasks, a second baseline measurement was carried out in which the participant watched another nature video. These baseline measures were not further analysed, as we were primarily interested in the effect of the manipulated task conditions.

Tasks and Instruments

Participants administered four tasks differing in terms of content topic (math versus non-math) and difficulty level (easy versus difficult): (1) easy math task, (2) difficult math task, (3) easy non-math task, and (4) difficult non-math task. Each task started with instructions and five practice trials, after which the instructions were repeated, and 35 experimental trials were presented. Each trial consisted of two stimuli presented on opposite sides of the screen until response. After a response was given using the arrow keys, an interstimulus interval with a fixation mark was presented for 1400 microseconds. Measures of reaction times and accuracies for each trial were recorded by the program PsychoPy.

The math task was a fraction comparison task in which participants verified which of two fractions was the largest (see an example of math items in Figure 2). Various indicators, of which multiple evidence indicates that these indicators impact the difficulty level of fraction comparisons, were used to create an easy and difficult condition of this task: natural number bias, distance effect, gap thinking, benchmarking, common components, and the number of digits (DeWolf & Vosniadou, 2011; Faulkenberry & Pierce, 2011; Vamvakoussi et al., 2012). The non-math task was a colour comparison task, in which participants had to decide on which side of the screen the mixing of colours resulted in the darkest colour (see an example of non-math items in Figure 1). This task was developed because of the similarity to the fraction comparison task, which is a two-alternatives with forced-choice verification, in which four elements must be processed and combined in a decision. Similarly as for the math task, six indicators were used to systematically manipulate an easy and a difficult version of the non-task: distance effect, exactness brightness, benchmarking, lightest/darkest, common components, and the need to mix. Although we aimed for the highest possible degree of equality, these tasks differed since the fraction comparison task requires semantic processing and the colour comparison task is a perceptual comparison task. Based on some pilots, the difficulty of this alternative non-math task was adjusted to match performance with the math task. Elaborated information about the development of the tasks and indicators can be found in the [Supplementary Materials](#) (see Figure S1 and Figure S2), and the tasks are available on the Open Science Framework (Demedts et al., 2023a).

After each task, participants had to rate two questions on a scale from 1 to 10. The first question was: "How difficult did you find the exercises?", to determine the perceived difficulty. The scale ranged from not difficult (1) to very difficult (10). Second, participants were asked to answer the following question: "How anxious were you while doing the exercises?". This item probed the state (math) anxiety, ranging from not anxious (1) to very anxious (10).

Physiological Measurements

Physiological data were measured by two devices developed by the research and development hub imec, see Figure 2. The first device was a chest patch (represented on the right side of Figure 2) recording an electrocardiograph at a sampling rate of 256 Hz, which reported information about HR and HRV. Second, a wrist-worn wearable (represented on the left side of Figure 2) registered GSR at a sampling rate of 256 Hz with a high dynamic range (0.05-20 μ S) at the lower side of the wrist and ST at the upper side of the wrist at a frequency of 1 Hz. This wearable was worn on the non-dominant hand and the chest patch was placed on the left chest. From the four recorded physiological signals, a total of 20 features were extracted. Redundant features were removed based on intercorrelations ($\max r = 0.7$) of features from the same signal (Smets et al., 2018; Wijsman et al., 2011). This resulted in a reduced feature set for further analysis of 9 features (2 GSR features, 3 ST features, 1 HR feature, and 3 HRV features; McCraty et al., 2009). Additionally, a quality indicator was calculated automatically by an algorithm that took anomalies into account to check the quality of the data measures. The data were averaged based on ticks for the different conditions.

Figure 2

Wearable Devices to Measure Physiological Responses



Analysis

For the physiological data, the quality indicator (ranging from 0 to 1) was checked before analysing the data. Values lower than .80 were deleted (for a more detailed explanation of this procedure, see Smets et al., 2018), because this indicates low quality of the data which could be caused by bad connection due to incorrect sensor attachment, too dry skin, chest hair (ECG patches), etc. Unfortunately, 9 participants had to be removed for the analyses on HR(V). As a result, the analyses for HR(V) data are only based on 35 participants.

To account for the repeated measurements and the large individual differences of physiological measurements, we opted to use multilevel analysis techniques, linear mixed models (LMM) with Restricted Maximum Likelihood estimation. All LMM analyses were performed in R using the lme4 package (Bates et al., 2015) and supplementary pairwise comparisons were carried out with the Multcomp package (Hothorn et al., 2008), using Holm correction. All numerical variables were centered and standardized (Lorah, 2018).

Before addressing the research questions, a manipulation check was performed by comparing the performances (accuracies and reaction times) and self-reported difficulty for the four task conditions, using a LMM and pairwise comparisons. We expected differences in terms of accuracies, reaction times and, self-reported difficulty between the easy and difficult task of the same topic. To be more precise, we assumed the difficult task versions to be administered less accurately and slower and to be perceived as more difficult compared to their easy version. For tasks of the same difficulty level from another topic, we expected these tasks to be similar in terms of accuracies and perceived difficulty. Since the processing for these tasks is distinct, we expected differences regarding reaction times in that math tasks require longer reaction times.

The first research question addressed whether the tasks resulted in differences in self-reported anxiety and physiological measures. Given this research question, we investigated the mean differences of these measures between the four task conditions. We used a univariate LMM for each outcome variable, and included topic and difficulty level, as fixed effects and additionally controlled for order effects. Multiple comparisons were conducted to identify differences between the four tasks.

The second research question investigated whether self-reported MA could be assessed through physiological measures, therefore this research question will be answered only with data from the math tasks. We first examined the zero-order repeated measures correlations between the self-reported MA and the physiological measures. To identify the features that can indicate self-reported MA, a separate multilevel analysis was carried out for each physiological signal (ST, HR, HRV, and GSR). These separate models provide a perception about how well these signals can indi-

cate self-reported MA. Features with a Variance Inflation Factor (VIF) higher than 10 were removed, due to high multicollinearity. Next, the features across signals that are significantly explanatory for self-reported MA are merged in a combined multilevel model. This combined model explains how well self-reported MA can be measured when combining physiological data.

Results

Manipulation Check

Before turning to the research questions, we checked whether the manipulation regarding the difficulty level of the tasks functionated properly. Table 1 summarizes the descriptive data for the four tasks, for the outcome variables accuracies, reaction times, and perceived difficulty. Furthermore, in Table 2 and Figure 3, results of multilevel analyses are depicted for each outcome variable to check the effect of the manipulation statistically. No significant order effects were found for these outcome variables.

Table 1

Descriptive Statistics for Accuracies, Reaction Times, and Perceived Difficulty

Topic	Difficulty level											
	Easy		Difficult		Easy		Difficult		Easy		Difficult	
	Accuracies				Reaction time				Perceived difficulty			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Math	32.70	4.10	23.09	4.59	1.63	0.91	5.75	5.68	4.23	2.36	8.36	1.31
Non-math	34.82	0.45	23.05	5.06	0.54	0.24	2.05	1.06	1.43	0.70	8.18	1.98

Table 2

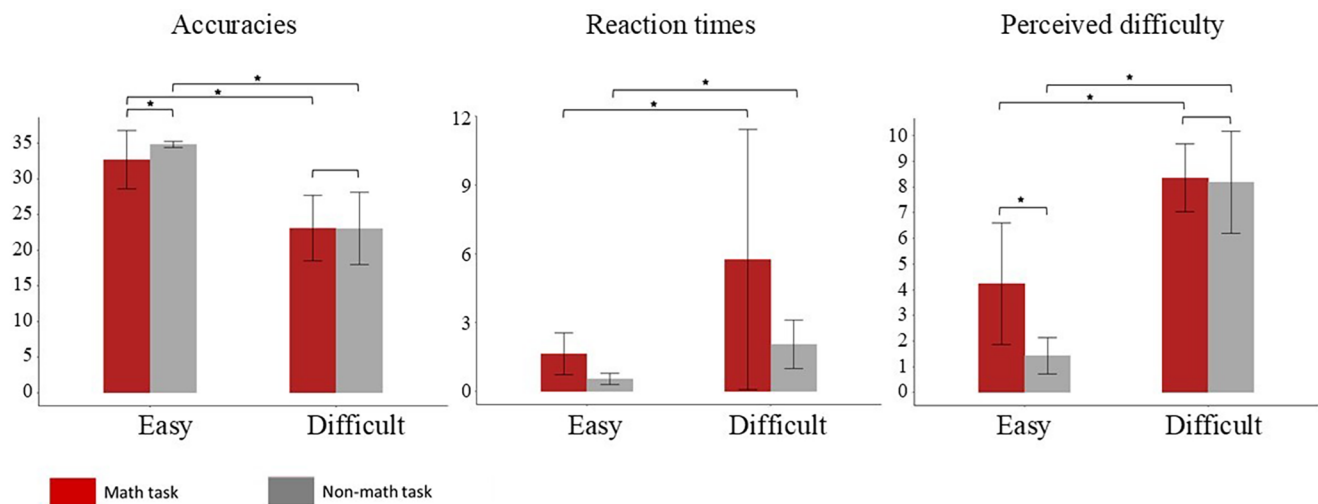
Results for the Multilevel Analyses of Accuracies, Reaction Times, and Perceived Difficulty

Fixed Effects	Accuracies				Reaction times				Perceived difficulty			
	<i>Est (SE)</i>	<i>t</i>	<i>p</i>	η_p^2	<i>Est (SE)</i>	<i>t</i>	<i>p</i>	η_p^2	<i>Est (SE)</i>	<i>t</i>	<i>p</i>	η_p^2
Intercept	-0.90 (0.21)	-4.25	< .001***		0.51 (0.29)	1.74	.08 [†]		0.86 (0.18)	4.84	< .001***	
Topic	-0.01 (0.12)	-0.06	.95	0.00	-1.06 (0.16)	-6.43	< .001***	0.24	-0.05 (0.09)	-0.58	.57	0.00
Difficulty	1.43 (0.12)	12.05	< .001***	0.53	-1.18 (0.16)	-7.18	< .001***	0.29	-1.23 (0.09)	-13.04	< .001***	0.57
Order	0.01 (0.04)	0.37	.71	0.00	0.03 (0.05)	0.54	.59	0.00	-0.03 (0.03)	-0.90	.37	0.01
General anxiety	0.01 (0.01)	0.51	.61	0.01	0.04 (0.02)	2.24	.04*	0.11	0.01 (0.01)	1.02	.32	0.02
Topic x Difficulty	0.32 (0.17)	1.92	.06 [†]	0.03	0.74 (0.23)	3.21	.002**	0.07	-0.78 (0.13)	-5.85	< .001***	0.21
Random Effects	Variance		SD		Variance		SD		Variance		SD	
Within subjects variance	0.31		0.56		0.59		0.77		0.19		0.44	
Between subjects variance	0.05		0.22		0.08		0.29		0.06		0.25	
Model Fit												
R_1^2	0.65				0.34				0.74			
R^2	0.69				0.42				0.81			
AIC	348.11				457.36				284.01			

[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Figure 3

Visualisation of the Manipulation Check per Condition



For the accuracies, a main effect of difficulty indicated better performance for the easy tasks than for the difficult tasks. Further, there was a (nearly) significant interaction between topic and difficulty. In line with our hypotheses, pairwise comparisons showed that difficult tasks were solved less accurately than their easy version, both for math ($\beta = -1.43$, $SE = 0.12$, $p < .001$) and non-math tasks ($\beta = -1.75$, $SE = 0.12$, $p < .001$). Moreover, there were no significant differences regarding accuracies for the two difficult tasks ($\beta = -0.01$, $SE = 0.12$, $p = .95$). Considering the easy tasks, however, results indicated that the easy math task was solved less accurately than the easy non-math task ($\beta = -0.32$, $SE = 0.12$, $p = .02$).

The LMM for reaction times showed main effects for difficulty and topic and an interaction. Pairwise comparisons showed an increase in reaction times for difficult tasks, both for math ($\beta = 1.18$, $SE = 0.16$, $p < .001$) and non-math tasks ($\beta = 0.43$, $SE = 0.16$, $p = .02$). The comparison of the reaction times between the two tasks of the same difficulty level but another topic (math vs non-math) is less relevant, given that the math task requires semantic processing and the non-math task is a perceptual comparison task.

In view of perceived difficulty, a main effect of difficulty and interaction was observed. Pairwise comparisons indicated that the difficult tasks were perceived as more difficult compared to the easy ones, for both math ($\beta = 1.23$, $SE = 0.09$, $p < .001$) and non-math tasks ($\beta = 2.00$, $SE = 0.09$, $p < .001$). As expected, there was no significant difference in the perceived difficulty of the two difficult tasks ($\beta = -0.05$, $SE = 0.09$, $p = .57$), but the easy math task was, on average, perceived as more difficult compared to the easy non-math task ($\beta = 0.83$, $SE = 0.09$, $p < .001$).

In sum, the results confirmed that the difficult tasks were resolved slower and less accurately and were perceived as more difficult by the participants compared to the easy versions of the same task. However, a (nearly) significant interaction was found for all outcome variables. For tasks of the same difficulty level but different in topic, it was observed that only the two difficult tasks were administered equally correctly and were perceived equally difficult by the participants. This was not the case for the two easy tasks, as the easy non-math task was completed slightly more accurately and was perceived as being slightly easier compared to the easy math task. We can therefore conclude that the manipulation did not fully worked out as expected. As a result, for the following research questions, when comparing math and non-math tasks, we will only compare the difficult tasks.

Differences in Measures of Anxiety in Easy and Difficult Math and Non-Math Tasks

Through the first research question, we want to gain insight into differences in measures of state anxiety, namely measured through self-reports and physiological measures, for four tasks with different difficulty and topic. Table 3 summarizes the descriptive data for the four tasks, for the self-reported anxiety and all physiological measures.

Table 3

Descriptive Statistics for Self-Reported Anxiety and Physiological Features

Topic	Difficulty level																			
	Easy		Difficult		Easy		Difficult		Easy		Difficult		Easy		Difficult		Easy		Difficult	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
	Self-reported anxiety				GSR: skin conductance level				GSR: skin conductance phasic				ST: mean				ST: standard deviation			
Math	3.82	2.20	5.30	2.40	0.24	0.48	0.27	0.58	4.45e-3	1.15e-2	7.18e-3	1.70e-2	29.30	2.04	29.26	2.04	0.04	0.03	0.08	0.06
Non-math	1.80	1.19	4.19	2.35	0.25	0.52	0.25	0.54	5.92e-3	1.99e-2	7.72e-3	2.42e-2	29.31	2.01	29.29	2.05	0.03	0.02	0.05	0.02
	ST: slope				HR: mean				HRV: standard deviation of RR intervals				HRV: low frequency divided by high frequency				HRV: heart coherence ratio			
Math	5.41e-4	1.27e-3	6.81e-4	1.35e-3	84.61	11.66	85.36	10.66	64.14	43.08	70.16	41.21	1.74	2.26	2.16	3.44	0.48	0.54	0.34	0.27
Non-math	7.34e-4	1.15e-3	4.34e-4	1.22e-3	83.23	11.84	83.25	11.33	60.13	42.29	62.55	37.76	1.81	2.02	1.85	1.99	0.33	0.21	0.36	0.27

Self-Reported Approach

Results of the self-reported anxiety, displayed in Table 4 and Figure 4, indicated that participants' self-reported state anxiety differed based on the topic and the difficulty of the task. There appeared no significant effect of order. A main effect of difficulty and topic was found. Pairwise comparisons showed that participants reported to experience more feelings of anxiety during the difficult tasks compared to their easy versions, for both math ($\beta = 0.61, SE = 0.13, p < .001$) and non-math tasks ($\beta = 0.92, SE = 0.13, p < .001$). For the difficult math task state anxiety levels were scored higher compared to the difficult non-math task ($\beta = 0.51, SE = 0.12, p < .001$).

Table 4

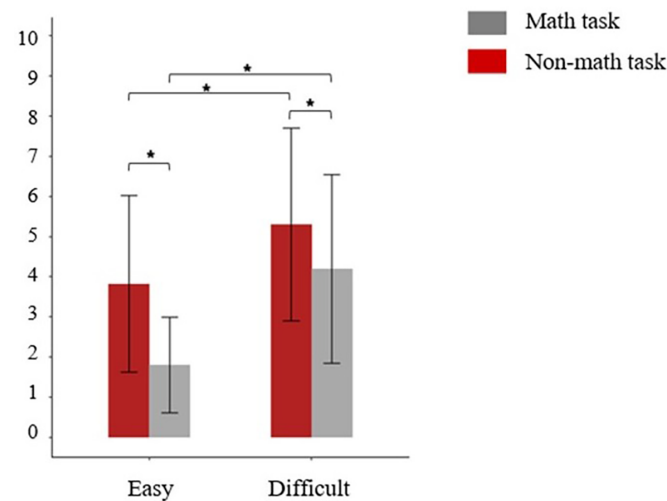
Results for the Multilevel Analyses of Participants' Self-Reported Math Anxiety

Fixed Effects	Self-reported anxiety				
	Est (SE)	t	p	η_p^2	
Intercept	0.30 (0.28)	1.05	.30		
Topic	-0.67 (0.09)	-7.52	< .001***	0.30	
Difficulty	-0.77 (0.09)	-8.58	< .001***	0.36	
General anxiety	0.04 (0.02)	1.89	.07 [†]	0.08	
Order	0.02 (0.04)	0.38	.70	0.00	
Random Effects	Variance			SD	
Within subjects variance	0.35			0.59	
Between subjects variance	0.38			0.61	
Model Fit					
R_1^2			0.29		
R^2			0.66		
AIC			415.67		

[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Figure 4

Visualisation of the Self-Reported Anxiety per Condition



Physiological Approach: Skin Temperature, Heart Rate (Variability) and Galvanic Skin Response

For ST ($N = 44$), a separate LMM was conducted for the mean, the standard deviation, and the slope. A main effect of order was found for the mean ST, indicating an increasing temperature throughout the experiment. No main effects were found for the mean and slope of the ST. For the standard deviation of the ST, a main effect of difficulty and topic was found. Pairwise comparisons between the tasks revealed a higher standard deviation of ST for the difficult compared to the easy version, for both math ($\beta = 0.93$, $SE = 0.18$, $p < .001$) and non-math ($\beta = 0.40$, $SE = 0.18$, $p = .07$) tasks. Further, a higher standard deviation of the ST was found for the difficult math task compared to the difficult non-math task ($\beta = 0.80$, $SE = 0.18$, $p < .001$).

For both the tonic and phasic GSR component ($N = 44$), no main effect of topic, difficulty, or order was found.

For the mean HR ($N = 35$), a main effect of topic and order was observed. No effect of difficulty was found. Pairwise comparisons did not result in any significant differences between the tasks.

Three features of HRV ($N = 35$) were analysed, more specifically the standard deviation of RR intervals, the low frequency divided by the high frequency, and the heart coherence ratio. Only for the standard deviation of RR intervals, a main effect for order was observed and a nearly significant main effect for topic and difficulty. Pairwise comparisons revealed only a significant difference between the easy non-math condition and the difficult math condition ($\beta = 0.26$, $SE = 0.09$, $p = .03$).

Together, regarding the first research question, it can be concluded that different tasks (varying in topic and difficulty) result in differences in self-reported anxiety, and to a limited extent in physiological responses. Self-reported state anxiety was higher for: (1) the difficult compared to the easy math task and (2) the difficult math task compared to the difficult non-math task. However, these self-reported differences in anxiety are not (or barely) detected by the physiological measures, except for some small differences between conditions for the standard deviation of the ST and the standard deviation of RR intervals (see [Supplementary Materials, Table S1, Table S2, and Table S3](#)). Overall, at a group level, we can conclude that increasing difficulty and changing topic does not result in significant differences in physiological measures.

Physiological Measures as Indicators for Self-Reported Math Anxiety

Since the purpose of the second research question was to identify physiological indicators for MA, we excluded the two non-math conditions from this analysis. To provide a first idea about the relatedness between the self-reported state-MA

and the various physiological features, a repeated measures correlation matrix is depicted in the [Supplementary Materials](#) (see Table S4). Results showed positive correlations between self-reported state-MA and the standard deviation of the ST ($r_m = .37, p = .01$), the standard deviation of the RR intervals ($r_m = .35, p = .04$), and a negative correlation with the heart coherence ratio ($r_m = -.38, p = .03$). Multilevel analyses were applied to further investigate these preliminary results of the repeated measures correlations.

As a first step, all features of each physiological signal were considered in a separate LMM. The VIF scores were low ($VIF < 3$), suggesting that there was no risk of multicollinearity in these analyses. The physiological measures that resulted to be significant predictors of the self-reported MA were combined in one multilevel model: the phasic GSR component, the standard deviation of the ST, and the heart coherence ratio as a measure of HRV.

However, results of this combined LMM indicated that the standard deviation of the ST was not a significant predictor of the self-reported MA, so this predictor was excluded. The results of the final LMM (see last column of Table 5) showed that higher self-reported MA was associated with an increased phasic GSR and a decreased heart coherence ratio, with small effect sizes. Together, these two physiological features explained 8% of the variance in self-reported MA. This model had a better fit than the null model, $\chi^2(2, N = 34) = 11.06, p < .01$.

Table 5

Results for the Multilevel Analyses of Participants' Self-Reported Math Anxiety

Fixed Effects	Null model				Model 1				Model 2			
	<i>b</i> (SE)	<i>t</i>	<i>p</i>	η_p^2	<i>b</i> (SE)	<i>t</i>	<i>p</i>	η_p^2	<i>b</i> (SE)	<i>t</i>	<i>p</i>	η_p^2
Intercept	0.00 (0.16)	0.01	1	–	0.00 (0.14)	0.00	1	–	0.00 (0.14)	0.00	1	–
GSR: skin conductance phasic	–	–	–	–	0.34 (0.13)	2.68	.01**	.11	0.35 (0.13)	2.76	.01**	.12
ST: standard deviation	–	–	–	–	0.02 (0.10)	0.21	.83	.00	–	–	–	–
ECG: heart coherence ratio	–	–	–	–	-0.21 (0.10)	-2.06	.04*	.08	-0.21 (0.10)	-2.11	.04*	.08
Random Effects												
	Variance		SD		Variance		SD		Variance		SD	
Within subjects variance	0.36		0.60		0.33		0.58		0.33		0.58	
Between subjects variance	0.65		0.80		0.53		0.73		0.52		0.72	
Model Fit												
R_1^2	0				0.07				0.08			
R^2	0				0.15				0.16			
AIC	175.62				170.52				168.55			

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

In sum, with regard to the second research question, these results suggest that some physiological features (i.e., phasic skin conductance and heart coherence ratio) can explain a part of the variance of the self-reported MA. However, the size of these effects is rather small. As a result, most of the variance in self-reported MA cannot be explained by these physiological measures.

Discussion

This study investigated to what extent self-reported state-MA can be measured through physiological measures. For this purpose, we manipulated the difficulty level of a math task to encompass a wide range of state-MA. Additionally, we designed a parallel easy and difficult non-math task, to investigate whether physiological signals are a response to being

confronted with the subject of mathematics or are just a reaction towards the difficulty level of a task. We administered the tasks in a lab setting to control for as many as possible confounding variables.

Manipulation Check of Performances and Perceived Difficulty

In line with our hypotheses, we observed significant differences for accuracies, reaction times, and perceived difficulty between the easy and difficult versions of both the math and the non-math tasks. This implies that the difficult versions were resolved less accurately accompanied by longer reaction times and were perceived more difficult, compared to the easy versions of the same task. As expected, no significant differences between the two difficult tasks were identified, indicating that both tasks were perceived as equally difficult and were solved equally accurately. By contrast, this was not the case for the two easy tasks, since the easy math task was perceived as more difficult and was solved less accurately than the easy non-math task.

These results only partially confirm the effectiveness of the manipulation of the difficulty level. Despite that the fractions included in the easy math task were of elementary school level, individuals apparently found them difficult. This finding is in line with research that revealed continuing difficulty with this topic (DeWolf & Vosniadou, 2011) and might be a possible explanation for the experienced difficulty of this easy math task. Another possible explanation relates to individual differences in experiencing something as difficult (Doz et al., 2023), causing students to perceive the easy math task as more difficult than the easy non-mathematics task. Future work may include defining difficulty or request participants to report their unique classification criteria.

Measures of State Anxiety in Math and Non-Math Tasks

The first research question investigated whether the different tasks resulted in differences in terms of state anxiety, measured through self-reports (subjective measure) and physiological data (objective measure). Regarding self-reported anxiety, the results of this empirical study show that both the topic and difficulty level of the tasks influenced feelings of anxiety.

The manipulation check confirmed that the math tasks could be distinguished based on their difficulty level since the easy task was found to be resolved faster and more accurately and was rated to be less difficult compared to the difficult version. This result suggested that the difficult math task triggered more MA, which is in line with studies pointing to the ‘anxiety-complexity effect’ (Trezise & Reeve, 2018). Although, some previous studies also provided evidence that MA already affects very basic mathematic processing (Maloney et al., 2010, 2011). Hence, it could conceivably be hypothesised that MA already impacts simple math tasks, however, task difficulty and MA are strongly related causing MA to increase with more difficult math tasks.

Furthermore, it must be noted that the self-reported (math) anxiety was relatively low in this study since in both difficult conditions only an average anxiety rate was reported (i.e., 5.30 and 4.19 on a scale from 1 to 10 for math and non-math tasks, respectively). One reason for this could be that there was no immediate effect or consequence for the participant when performing low-level on this experiment. Therefore, it might be worthwhile to investigate this research question more thorough in high-stakes and more ecologically valid situations in further studies (as also been suggested by Strohmaier et al., 2020). Another possible explanation for these relatively low levels of self-reported anxiety may be the difficulties associated with assessing affective states retrospectively, as described in the introduction (Dowker et al., 2016).

Despite these low self-reported anxiety levels, it is striking that we observed a significant difference between the two difficult tasks, indicating that the math task aroused more feelings of anxiety than the non-math task. In addition, the manipulation check demonstrated that these tasks were equally difficult in terms of accuracies and, were additionally perceived as such by the participants. So, the differences in self-reported anxiety cannot be due to the fact that a particular task was more difficult or was perceived to be so. This finding suggests the existence of a domain-specific MA construct, indicating that mathematical situations specifically trigger feelings of anxiety.

Concerning the physiological measures, potential differences between conditions were only observed for two features. The standard deviation of the ST was higher in the difficult tasks compared to the easy ones and was higher in the difficult math task compared to the difficult non-math task. The standard deviation of RR intervals was

significantly higher in the difficult math task compared to the easy math task and the difficult non-math task. For all other features, there were no significant differences between conditions. This finding is consistent with results from [Hunt and colleagues \(2017\)](#) who could only observe differences in physiological measures between a baseline condition and a task condition, but not between two task conditions differing in difficulty. Moreover, this finding is in line with the observation that self-reported anxiety was rather low for the provided tasks. A possible explanation might thus also in this case be that the experiment did not induce a very strong feeling of anxiety, resulting in differences in physiological responses that were not sensitive enough to be registered. However, it is also possible that the challenges associated with using physiological measures prevented us from finding significant results ([Horvers et al., 2021](#)). In addition, further research with a stronger focus on the way these physiological responses are interpreted by individuals may provide additional insights.

In sum, significant differences in terms of self-reported anxiety were observed between the four tasks conditions. By contrast, only two differences between these tasks were found for the physiological measures. It can be concluded that differences in self-reported anxiety are (almost) undetected by the physiological measures (in line with [Strohmaier et al., 2020](#)).

Physiological Measures as Indicators of Self-Reported Math Anxiety

The second research question examined to what extent the variability in self-reported MA, through an easy and difficult math task, can be assessed through physiological measures. Our data revealed no strong associations between participants' self-reported feelings of state MA and physiological measures. These results seem to be consistent with other research also reporting no significant correlations ([Strohmaier et al., 2020](#)) or only in very limited timeframes ([Qu et al., 2020](#)). From the multilevel analyses, it could be concluded that higher values in phasic GSR and lower values of heart coherence ratio are significantly indicative for self-reported MA. Although two physiological features are shown to be predictive for the self-reported feelings of MA, a note of caution is due here since only a limited amount of variance in self-reported MA is explained by these two physiological features.

There are several possible explanations for these results. One possible explanation could be that these two measurement approaches assess two distinct facets of the same construct, as also suggested by [Strohmaier and colleagues \(2020\)](#). Alternatively, it is also possible that the relationship between self-reported anxiety and physiological responses only occurs at high levels of anxiety, thus involving a threshold linear relationship. It can therefore be assumed that at this moment, physiological responses should not yet be used as a primary method for mapping MA ([Harley et al., 2015](#)). To further develop a full picture of the possibility to use individuals' physiological responses as a measure of their level of MA, further research is needed.

Limitations

When interpreting the results and drawing conclusions, it is important to consider the limitations that are inherently connected to the present study. First, we should acknowledge the relative small number of participants in this study. The limited number of studies relying on physiological indices complicates determining an appropriate sample size. Nevertheless, we would like to point out that the number of participants in the current study is comparable to previous studies that have included physiological data (e.g., [Singh et al., 2019](#); [Qu et al., 2020](#)). Moreover, the applied within-subject design with a randomized two-by-two factorial design adds to the power of this study (i.e., 44 level two units, 176 level one units). Since we only examine fixed effects, the number of clusters in this study is proved sufficient to lead to good and unbiased estimates ([Maas & Hox, 2004](#)). Second, using physiological measures comes with some limitations. For example, several features were extracted from the physiological responses, but of course, it still concerns a selection which could be optimised. Furthermore, physiological measures are missing for some participants due to poor connectivity of the devices resulting into low data quality. However, relying on a multimodal approach is seen as a relevant avenue to gain a deeper understanding of how feelings of math anxiety manifest in individuals. Third, we must indicate the limitations of the used tasks. The predetermined manipulation did not prove to be completely successful, leaving us to restrict the analysis to the comparisons that could be reliably performed. Possibly different processing strategies are used in the different tasks and non-math tasks contained no academic content, however the current data

collection does not allow to uncover the underlying rationale. Regarding the math tasks, we recognize that these tasks are not comparable to mathematical contexts students face in everyday life, which might have influenced the occurrence of authentic feelings of MA.

Given these limitations, there is a need for replication studies with larger samples in more authentic settings using an alternative control task (after extensive piloting). Eye-tracking and think-a-loud methods can be added to gain more insight into how students perceive and solve the tasks, and to further explore the use of physiological responses to validate our findings.

Conclusion

To conclude, the current study has shown that two equally difficult tasks resulted in differences in terms of self-reported anxiety, indicating a specific anxiety associated with mathematical tasks. However, it was somewhat surprising that almost no differences in physiological measures were noted between those tasks. Moreover, this study indicated that some physiological measures (i.e., phasic GSR and heart coherence ratio) are significant predictors for individuals' self-reported MA. Further research, which considers the abovementioned methodological limitations, is necessary to provide more insights into the use of the physiological measures in the research field of MA.

Funding: This research was funded by the Research Foundation – Flanders under Grant G068520N.

Acknowledgments: The authors have no additional (i.e., non-financial) support to report.

Competing Interests: All authors wish to confirm that there are no conflicts of interest related to this publication.

Ethics Statement: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and approved by the Social and Societal Ethics Committee (SMEC) of the KU Leuven (G-2020-2509).

Data Availability: The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Supplementary Materials

The Supplementary Materials include a detailed description of the designed tasks, the results of the multilevel analyses for Research Question 1 for the different physiological responses and a correlation matrix for the physiological responses and self-reported math anxiety (see Demedts et al., 2023b).

On the Open Science Framework, we have uploaded the full set of tasks so that the tasks can be reused (see Demedts et al., 2023a).

Index of Supplementary Materials

Demedts, F., Cornelis, J., Reynvoet, B., Sasanguie, D., & Depaepe, F. (2023a). *Measuring math anxiety through self-reports and physiological data: Experimental tasks* [Task set]. OSF. <https://osf.io/6jw2t/>

Demedts, F., Cornelis, J., Reynvoet, B., Sasanguie, D., & Depaepe, F. (2023b). *Supplementary materials to "Measuring math anxiety through self-reports and physiological data"* [Additional information]. PsychOpen GOLD. <https://doi.org/10.23668/psycharchives.13492>

References

- Artemenko, C., Daroczy, G., & Nuerk, H.-C. (2015). Neural correlates of math anxiety – An overview and implications. *Frontiers in Psychology*, 6, Article 1333. <https://doi.org/10.3389/fpsyg.2015.01333>
- Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science*, 11(5), 181–185. <https://doi.org/10.1111/1467-8721.00196>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>

- Bieg, M., Goetz, T., & Lipnevich, A. A. (2014). What students think they feel differs from what they really feel – Academic self-concept moderates the discrepancy between students' trait and state emotional self-reports. *PLoS ONE*, *9*(3), Article e92563. <https://doi.org/10.1371/journal.pone.0092563>
- Boucsein, W. (2012). *Electrodermal activity*. Springer.
- Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2015). *A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments* (Technical Report, 2nd version: Selective Attention & Awareness Laboratory [SAAL], Behavioural Brain Sciences Centre, University of Birmingham, United Kingdom). <https://www.birmingham.ac.uk/documents/college-les/psych/saal/guide-electrodermal-activity.pdf>
- Carey, E., Hill, F., Devine, A., & Szűcs, D. (2016). The chicken or the egg? The direction of the relationship between mathematics anxiety and mathematics performance. *Frontiers in Psychology*, *6*, Article 1987. <https://doi.org/10.3389/fpsyg.2015.01987>
- Cargnelutti, E., Tomasetto, C., & Passolunghi, M. C. (2017). How is anxiety related to math performance in young students? A longitudinal study of Grade 2 to Grade 3 children. *Cognition and Emotion*, *31*(4), 755–764. <https://doi.org/10.1080/02699931.2016.1147421>
- Cipora, K., Artemenko, C., & Nuerk, H.-C. (2019). Different ways to measure math anxiety. In I. Mammarella, S. Caviola, & A. Dowker (Eds.), *Mathematics anxiety: What is known, and what is still missing* (pp. 20–41). Routledge.
- Conlon, R. A., Hicks, A., Barroso, C., & Ganley, C. M. (2021). The effect of the timing of math anxiety measurement on math outcomes. *Learning and Individual Differences*, *86*, Article 101962. <https://doi.org/10.1016/j.lindif.2020.101962>
- Dew, K. H., Galassi, J. P., & Galassi, M. D. (1984). Math anxiety: Relation with situational test anxiety, performance, physiological arousal, and math avoidance behavior. *Journal of Counseling Psychology*, *31*(4), 580–583. <https://doi.org/10.1037/0022-0167.31.4.580>
- DeWolf, M., & Vosniadou, S. (2011). The whole number bias in fraction magnitude comparisons with adults. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1751–1756). Cognitive Science Society.
- Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology*, *7*, Article 1333. <https://doi.org/10.3389/fpsyg.2016.00508>
- Doz, E., Cuder, A., Pellizzoni, S., Carretti, B., & Passolunghi, M. C. (2023). Arithmetic word problem-solving and math anxiety: The role of perceived difficulty and gender. *Journal of Cognition and Development*, *24*(4), 598–616. <https://doi.org/10.1080/15248372.2023.2186692>
- Faulkenberry, T. J., & Pierce, B. H. (2011). Mental representations in fraction comparison holistic versus component-based strategies. *Experimental Psychology*, *58*(6), 480–489. <https://doi.org/10.1027/1618-3169/a000116>
- Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., & Tsiknakis, M. (2022). Review on psychological stress detection using biosignals. *IEEE Transactions on Affective Computing*, *13*(1), 440–460. <https://doi.org/10.1109/TAFFC.2019.2927337>
- Halme, H., Trezise, K., Hannula-Sormunen, M. M., & McMullen, J. (2022). Characterizing mathematics anxiety and its relation to performance in routine and adaptive tasks. *Journal of Numerical Cognition*, *8*(3), 414–429. <https://doi.org/10.5964/jnc.7675>
- Harley, J. M. (2016). Measuring emotions: A survey of cutting edge methodologies used in computer-based learning environment research. In S. Y. Tettegah & M. Gartmeier (Eds.), *Emotions, technology, design, and learning* (pp. 89–114). Elsevier.
- Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior*, *48*, 615–625. <https://doi.org/10.1016/j.chb.2015.02.013>
- Hill, F., Mammarella, I. C., Devine, A., Caviola, S., Passolunghi, M. C., & Szűcs, D. (2016). Maths anxiety in primary and secondary school students: Gender differences, developmental changes and anxiety specificity. *Learning and Individual Differences*, *48*, 45–53. <https://doi.org/10.1016/j.lindif.2016.02.006>
- Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The Abbreviated Math Anxiety Scale (AMAS): Construction, validity, and reliability. *Assessment*, *10*(2), 178–182. <https://doi.org/10.1177/1073191103010002008>
- Horvers, A., Tombeng, N., Bosse, T., Lazonder, A. W., & Molenaar, I. (2021). Detecting emotions through electrodermal activity in learning contexts: A systematic review. *Sensors*, *21*(23), Article 7869. <https://doi.org/10.3390/s21237869>
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363. <https://doi.org/10.1002/bimj.200810425>
- Hunt, T. E., Bhardwa, J., & Sheffield, D. (2017). Mental arithmetic performance, physiological reactivity and mathematics anxiety amongst U.K. primary school children. *Learning and Individual Differences*, *57*, 129–132. <https://doi.org/10.1016/j.lindif.2017.03.016>

- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421. <https://doi.org/10.1016/j.biopsycho.2010.03.010>
- Liu, W. C., & Wang, C. K. J. (2005). Academic self-concept: A cross-sectional study of grade and gender differences in a Singapore secondary school. *Asia Pacific Education Review*, *6*(1), 20–27. <https://doi.org/10.1007/BF03024964>
- Lorah, J. (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-Scale Assessments in Education*, *6*, Article 8. <https://doi.org/10.1186/s40536-018-0061-2>
- Maas, C. J. M., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica*, *58*(2), 127–137. <https://doi.org/10.1046/j.0039-0402.2003.00252.x>
- Maloney, E. A., Ansari, D., & Fugelsang, J. A. (2011). The effect of mathematics anxiety on the processing of numerical magnitude. *Quarterly Journal of Experimental Psychology*, *64*(1), 10–16. <https://doi.org/10.1080/17470218.2010.533278>
- Maloney, E. A., Risko, E. F., Ansari, D., & Fugelsang, J. (2010). Mathematics anxiety affects counting but not subitizing during visual enumeration. *Cognition*, *114*(2), 293–297. <https://doi.org/10.1016/j.cognition.2009.09.013>
- Mammarella, I. C., Caviola, S., & Dowker, A. (2019). *Mathematics anxiety what is known and what is still to be understood*. Routledge.
- Marsh, H. W., & O'Neill, R. (1984). Self Description Questionnaire III: The construct validity of multidimensional self-concept ratings by late adolescents. *Journal of Educational Measurement*, *21*(2), 153–174. <https://doi.org/10.1111/j.1745-3984.1984.tb00227.x>
- McCraty, R., Atkinson, M., Tomason, D. B. A., & Bradley, R. T. (2009). The coherent heart: Heart-brain interactions, psychophysiological coherence, and the emergence of system-wide order. *Integral Review*, *5*(2), 10–115.
- Mertens, J., Zane, E. R., Neumeyer, K., & Grossman, R. B. (2017). How anxious do you think I am? Relationship between state and trait anxiety in children with and without ASD during social tasks. *Journal of Autism and Developmental Disorders*, *47*(12), 3692–3703. <https://doi.org/10.1007/s10803-016-2979-y>
- Namkung, J. M., Peng, P., & Lin, X. (2019). The relation between mathematics anxiety and mathematics performance among school-aged students: A meta-analysis. *Review of Educational Research*, *89*(3), 459–496. <https://doi.org/10.3102/0034654319843494>
- Noroozi, O., Pijera-Díaz, H. J., Sobocinski, M., Dindar, M., Järvelä, S., & Kirchner, P. A. (2020). Multimodal data indicators for capturing cognitive, motivational, and emotional learning processes: A systematic literature review. *Education and Information Technologies*, *25*, 5499–5547. <https://doi.org/10.1007/s10639-020-10229-w>
- Núñez-Peña, M. I., & Suárez-Pellicioni, M. (2014). Less precise representation of numerical magnitude in high math-anxious individuals: An ERP study of the size and distance effects. *Biological Psychology*, *103*, 176–183. <https://doi.org/10.1016/j.biopsycho.2014.09.004>
- Orbach, L., Herzog, M., & Fritz, A. (2019). Relation of state- and trait-math anxiety to intelligence, math achievement and learning motivation. *Journal of Numerical Cognition*, *5*(3), 371–399. <https://doi.org/10.5964/jnc.v5i3.204>
- Orbach, L., Herzog, M., & Fritz, A. (2020). State- and trait-math anxiety and their relation to math performance in children: The role of core executive functions. *Cognition*, *200*, Article 104271. <https://doi.org/10.1016/j.cognition.2020.104271>
- Pekrun, R., & Bühner, M. (2014). Self-report measures of academic emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 561–579). Routledge.
- Poh, M.-Z., Swenson, N. C., & Picard, R. W. (2010). A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Transactions on Biomedical Engineering*, *57*(5), 1243–1252. <https://doi.org/10.1109/TBME.2009.2038487>
- Qu, Z., Chen, J., Li, B., Tan, J., Zhang, D., & Zhang, Y. (2020). Measurement of high-school students' trait math anxiety using neurophysiological recordings during math exam. *IEEE Access: Practical Innovations, Open Solutions*, *8*, 57460–57471. <https://doi.org/10.1109/ACCESS.2020.2982198>
- Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math anxiety: Past research, promising interventions, and a new interpretation framework. *Educational Psychologist*, *53*(3), 145–164. <https://doi.org/10.1080/00461520.2018.1447384>
- Richardson, F. C., & Suinn, R. M. (1972). The mathematics anxiety rating scale: Psychometric data. *Journal of Counseling Psychology*, *19*, 551–554. <https://doi.org/10.1037/h0033456>
- Roos, A.-L., Goetz, T., Voracek, M., Krannich, M., Bieg, M., Jarrell, A., & Pekrun, R. (2021). Test anxiety and physiological arousal: A systematic review and meta-analysis. *Educational Psychology Review*, *33*, 579–618. <https://doi.org/10.1007/s10648-020-09543-z>
- Singh, N., Aggarwal, Y., & Sinha, R. K. (2019). Heart rate variability analysis under varied task difficulties in mental arithmetic performance. *Health and Technology*, *9*(3), 343–353. <https://doi.org/10.1007/s12553-018-0272-0>

- Smets, E., Velazquez, E. R., Schiavone, G., Chakroun, I., D'Hondt, E., De Raedt, W., Cornelis, J., Janssens, O., Van Hoecke, S., Claes, S., Van Diest, I., & Van Hoof, C. (2018). Large-scale wearable data reveal digital phenotypes for daily-life stress detection. *NPJ Digital Medicine*, 1, Article 67. <https://doi.org/10.1038/s41746-018-0074-9>
- Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, 166(10), 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>
- Strohmaier, A. R., Schiepe-Tiska, A., & Reiss, K. M. (2020). A comparison of self-reports and electrodermal activity as indicators of mathematics state anxiety: An application of the control-value theory. *Frontline Learning Research*, 8(1), 16–32. <https://doi.org/10.14786/flr.v8i1.427>
- Suárez-Pellicioni, M., Núñez-Peña, M. I., & Colomé, A. (2013). Mathematical anxiety effects on simple arithmetic processing efficiency: An event-related potential study. *Biological Psychology*, 94(3), 517–526. <https://doi.org/10.1016/j.biopsycho.2013.09.012>
- Suárez-Pellicioni, M., Núñez-Peña, M. I., & Colomé, A. (2016). Math anxiety: A review of its cognitive consequences, psychophysiological correlates, and brain bases. *Cognitive, Affective & Behavioral Neuroscience*, 16(1), 3–22. <https://doi.org/10.3758/s13415-015-0370-7>
- Trezise, K., & Reeve, R. A. (2018). Patterns of anxiety in algebraic problem solving: A three-step latent variable analysis. *Learning and Individual Differences*, 66, 78–91. <https://doi.org/10.1016/j.lindif.2018.02.007>
- Vamvakoussi, X., Van Dooren, W., & Verschaffel, L. (2012). Naturally biased? In search for reaction time evidence for a natural number bias in adults. *The Journal of Mathematical Behavior*, 31(3), 344–355. <https://doi.org/10.1016/j.jmathb.2012.02.001>
- Wijsman, J., Grundlehner, B., Liu, H., Hermens, H., & Penders, J. (2011). Towards mental stress detection using wearable physiological sensors. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 33, 1798–1801. <https://doi.org/10.1109/IEMBS.2011.6090512>



Journal of Numerical Cognition (JNC) is an official journal of the Mathematical Cognition and Learning Society (MCLS).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.