


RESEARCH

Open Access



The S in STEM: gender differences in science anxiety and its relations with science test performance-related variables

Dmitri Rozgonjuk^{1*} , Karin Täht^{1,2}, Regina Soobard³, Moonika Teppo³ and Miia Rannikmäe^{3*}

Abstract

Background STEM education has experienced significant growth due to its pivotal role in innovation and economic development. While cognitive factors like prior knowledge are known predictors of STEM success, non-cognitive factors, including attitudes and demographics, also play vital roles. However, there is a notable scarcity of research focusing on the "S" in STEM—science—compared to extensive studies in fields like mathematics. This study aims to address this gap by exploring gender differences in science test performance and related attitudes, providing insights into this under-researched aspect of STEM education.

Results The effective sample comprised 1839 Estonian 12th-grade students who took a computer-assisted science test. The test consisted of tasks combining chemistry, physics, biology, and geography, and a post-test survey was also administered. Across the total sample, the results showed that test performance positively correlated with test-taking duration, effort, and test importance. Test performance was negatively correlated with perceived test difficulty. Interestingly, while general science anxiety was not associated with test performance, subject-specific anxiety, especially chemistry anxiety had a negative association with test performance. While there were no gender differences in test performance, female students scored consistently higher on all science anxiety measures, compared to male students. Furthermore, female students assessed the science test to be more difficult, and they also took more time to complete the test. The correlations in gender subsamples mirrored those observed in the total sample.

Conclusions The association between science test performance and test-related variables is nuanced: students might not necessarily have a "general" STEM anxiety but it may be associated with a specific subject. Moreover, the findings imply that although there are no gender differences in test performance, girls have a greater anxiety when it comes to natural sciences subjects. These findings indicate the need for investigating the origin of such anxieties, which do not seem to stem from aptitude.

Keywords Science education, Science anxiety, Test performance, Secondary education, Chemistry, Physics, Biology, Geography, Gender differences, Gender gap in STEM

Introduction

Over the past few decades, the emphasis on Science, Technology, Engineering, and Mathematics (STEM) education has increased substantially (Kayan-Fadlelmula et al., 2022; Li et al., 2020). This surge in emphasis is not without reason: STEM fields stand at the forefront of innovation, playing a pivotal role in driving economic growth and addressing global challenges (Widya et al., 2019). STEM encompasses an array of disciplines, each

*Correspondence:

Dmitri Rozgonjuk
research@rozgonjuk.com
Miia Rannikmäe
miia.rannikmae@ut.ee

¹ Institute of Psychology, University of Tartu, Tartu, Estonia

² Institute of Mathematics and Statistics, University of Tartu, Tartu, Estonia

³ Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

intricate in its own way. In students' STEM performance at school, both cognitive and non-cognitive factors contribute. Cognitive factors, including prior knowledge, spatial skills, abstract thinking, and logical reasoning have been identified as significant predictors of success in STEM fields (Andersen, 2014; Berkowitz & Stern, 2018). Among non-cognitive factors, attitudes towards STEM subjects and fields have also been shown to play a significant role (Potvin & Hasni, 2014; Potvin et al., 2020). Additionally, demographic factors, such as socio-economic status (Rozgonjuk et al., 2023), and individual characteristics, such as gender (Wang et al., 2023), can significantly influence both academic performance and development of STEM attitudes.

Although research exists on the interplay between academic achievement and attitudes in STEM—and gender differences in the domain—specific aspects, such as the association between science test performance and motivational variables like anxiety, still warrant further investigation. Furthermore, science can be decomposed into more specific subjects, including physics, chemistry, biology, and geography. In the present work, we aim to investigate the association between science achievement, test-related attitudes as well as general and subject-specific science anxiety, and gender differences in these links.

Literature review

Historically, gender differences in STEM fields, particularly in science, have been a focal point of educational and psychological research (Li et al., 2020). The so-called *gender-equality paradox* refers to a phenomenon observed in many societies where, despite significant efforts to achieve gender equality across various fields, a notable discrepancy persists in the representation of men and women in certain professions, particularly those related to STEM (Stoet & Geary, 2018; Tandrayen-Ragoobur & Gokulsing, 2022). Not only do gender disparities exist in participation and representation in these fields, but studies have also indicated differences in performance, attitudes, and experiences (Voyer & Voyer, 2014). In Estonia, the proportion of female STEM graduates is generally higher than in other OECD countries; however, the pay gap between men and women remains one of the largest among the surveyed countries (OECD, 2024). Furthermore, some evidence suggests an association between a larger pay gap and a higher proportion of women in STEM-related fields, possibly reflecting a higher pay gap in STEM jobs (Treialt, 2021). Of relevance, the average age of Estonian teachers (including in science) is one of the highest among OECD countries (OECD, 2019b, p. 84), and it is a major factor in teaching motivation (Täht et al., 2023).

Several factors have been proposed to contribute to the development of the gender gap in STEM, including social and cultural influences and stereotypes (Cheryan et al., 2017; Master et al., 2021; Verdugo-Castro et al., 2022), differences in career interests and priorities (So et al., 2022), socio-economic status (Early et al., 2020; Rozgonjuk et al., 2023), differences in personality traits (Anni et al., 2023; Hofmann et al., 2023; McKinney et al., 2021), and societal expectations regarding gender roles (Hägglund & Leuze, 2021; Schmitt et al., 2008).

Academic self-efficacy and anxiety are two psychological concepts that are often intertwined when it comes to shaping the attitudes towards academic subjects (Burns et al., 2021; McKinney et al., 2021; Rozgonjuk et al., 2020). Self-efficacy is the belief in one's own abilities (Bandura, 1997), whereas academic anxiety refers to feelings of fear, apprehension, or worry experienced in response to academic tasks, assessments, or performance expectations (Tobias & Weissbrod, 1980). Research has shown that self-efficacy correlates with better academic outcomes in general (Warren et al., 2020), as well as more specifically in science (Burns et al., 2021) and mathematics (Özcan & Eren Gümüş, 2019). Conversely, poorer academic performance in respective domains has been shown to be accompanied by higher anxiety in mathematics (Caviola et al., 2022; Guzmán et al., 2023; Namkung et al., 2019) and in science (Megreya et al., 2021).

According to the meta-analysis by Reilly et al. (2015), male students tend to have a higher achievement in both mathematics and science than female students in the U.S. On the other hand, recent research, relying on international studies, shows a more mixed picture: in general, while boys outperform girls in mathematics, girls outperformed boys in science in most of the participating OECD countries in the PISA 2018 survey (OECD, 2019a, p. 142). These findings suggest that gender disparities in STEM achievement are context-dependent, highlighting the importance of considering cultural, educational, and socio-economic factors when addressing gender gaps in education.

In studies that do not find gender differences in academic outcomes, girls tend to score lower on mathematics self-efficacy (Zander et al., 2020) and higher on mathematics anxiety measures (Vos et al., 2023). In another study, Megreya et al. (2021) found that female students had higher science anxiety. Relatedly, Cotner et al. (2020) found that female Norwegian biology students reported higher testing anxiety than male biology students; interestingly, anxiety did not predict the achievement of male students. Moreover, in a recent systematic-narrative literature review, Balducci (2023) has also investigated the association between the gender gap in mathematics and science and country-level gender

equality. Interestingly, while mathematics gender gap was not linked to country-level gender equality, larger gender differences in mathematics attitudes and anxiety are reported in more gender-equal countries. When it comes to science, Balducci (2023) found that girls display a lower science self-concept than boys in the vast majority of cases—even though, as mentioned above, boys did not necessarily outperform girls in terms of ability.

These differences are not without consequences. For instance, Sakellariou and Fang (2021) argue that early development of self-efficacy in mathematics and science is predictive of propensity to study STEM subjects. The authors also showed that reducing the STEM gender gap is effective in girls with above-average self-efficacy. It could be argued, therefore, that by reducing STEM-related anxiety (which would, conversely, likely result in improving STEM self-efficacy), more girls could consider a STEM career path. Reducing anxiety and improving self-efficacy do not solely depend on one's performance but can also be achieved through targeted interventions addressing anxiety sources, teaching methods, and learning-related activities (Zakariya, 2022).

Conceptual framework

In this study, we contextualize gender differences in science, extrapolated to STEM education in general within the framework of Social Cognitive Theory (SCT; Bussey & Bandura, 1999). SCT posits that individual beliefs, environmental influences, and behaviors dynamically interact to shape academic outcomes. Building upon this theory, we propose that gender disparities in STEM fields result from a complex interplay of personal factors (such as self-efficacy and anxiety), socio-cultural influences (including stereotypes and societal expectations), and environmental factors (such as educational experiences and opportunities). Transitioning to our specific focus, we delve into investigating the variance in personal factors, notably science anxiety. This choice is informed by our understanding that individual attitudes and emotions can significantly impact academic performance and career choices within STEM disciplines.

In addition, although initially derived from the mathematics domain, the insights from Baloglu and Kocak (2006) can be plausibly extended to science. They suggest that achievement-related attitudes and motivational factors—in the context of the present study, science anxiety—can be influenced by situational, dispositional, and environmental factors. Situational elements pertain directly to the domain in question (i.e., testing in the present study). In contrast, dispositional and environmental factors revolve around the inclination to cultivate stable attitudes and behaviors towards subjects and the external contexts associated with the domain, respectively.

Hence, science anxiety could be affected by previous testing experiences (situational), general tendencies to experience anxiety (dispositional), and societal expectations and stereotypes regarding science performance (environmental).

The framework by Baloglu and Kocak (2006) is relevant to our study in linking science anxiety to situational test-related variables to provide empirical evidence on these associations. Investigating other factors (i.e., dispositional and environmental) is not within the scope of the present study. It is important to note that although the science anxiety inventory items refer to anxiety experienced towards science beyond testing situations, we queried about these attitudes after the test was taken. Hence, it is possible that, on one hand, anxiety could impact test performance; on the other hand, test-related variables like test difficulty, test duration, and test appeal could influence self-reported science anxiety, which might, in turn, impact attitudes towards science.

Aims and hypotheses

While science represents a component of STEM, emphasis on examining achievements in science in relation to anxiety, and the gender differences in that link, have garnered less attention than areas like mathematics. The main aim of the present study is to explore the associations between and gender differences in science test performance, test-taking time, science anxiety, and test-related attitudes. It could be argued that students who have lower science anxiety are more motivated to perform well on a test (i.e., they find the test important to them), and are also willing to exert more effort (Glynn et al., 2009). This, in turn, could lead to better performance. There is evidence in literature supporting this logic in test-taking research. Lower test anxiety is related to higher motivation and better academic performance, and students with higher test anxiety have an increased potential to engage in academic self-handicapping behavior. In addition, one may argue that higher motivation to perform well can also lead students to spend more time and apply greater diligence to their test-taking. In other words, more time spent on test-taking should correlate with higher motivation, and this could lead to better outcomes. However, based on literature, one may also anticipate that girls report higher science-related anxiety in comparison to boys—despite similar academic outcomes. Given this reasoning, the following hypotheses are posed:

H1: *Science anxiety measures are negatively correlated with science test performance.* High levels of anxiety, particularly when specific to a subject like science, can hinder a student's ability to access and utilize accumulated knowledge (Mallow, 2006). Research has consistently shown that high anxiety levels can interfere with

learning and impair test performance (Caviola et al., 2022). In addition, the link between academic performance and anxiety is relatively well-established in other STEM domains, like mathematics (Caviola et al., 2022; Guzmán et al., 2023; Namkung et al., 2019), but relatively less researched in the science domain.

H2: *Motivational variables (test importance, effort) are positively correlated with test performance.* Test-taking motivation and effort are closely connected. Students with higher motivation levels are more likely to exert effort in preparing for and taking the test, in turn leading to improved results (Alhadabi & Karpinski, 2020). The significance of motivation is evident in situations where a student possesses the requisite knowledge but may underperform due to a lack of motivation or perceived relevance of the test to their future goals (Dökme et al., 2022).

H3: *Test-taking duration is positively correlated with test importance and effort, as well as test performance.* While one might intuitively assume that faster completion times reflect higher levels of proficiency, empirical evidence suggests that rapid guessing (i.e., shorter test-taking time) correlates negatively with test scores (Rios et al., 2022). On the other hand, more time taken to solve assignments is generally linked to better results also in mathematics and science (Silm et al., 2020).

H4. *There are no gender differences in science test performance but girls score higher on science anxiety measures.* This hypothesis is inspired by findings from STEM research, particularly in mathematics, where there are generally no gender differences in academic performance, but girls report higher levels of STEM-related anxiety (Foley et al., 2017; Guzmán et al., 2023).

H5. *The correlation between test performance and anxiety is stronger in girls than in boys.* In other words, anxiety plays a larger role in performance in girls. Previously, it has been shown that the link between mathematics anxiety and performance is stronger in girls than in boys (Devine et al., 2012; Dowker et al., 2016). Additionally, this hypothesis is inspired by the findings in literature that showed that despite having similar academic outcomes in STEM subjects, girls tend to opt for a STEM career less (Schmader, 2023; Sevilla & Snodgrass Rangel, 2023). The findings could provide an additional explanation to that phenomenon: it could be that STEM is associated with more anxiety for girls than for boys, consequently guiding the former away from a STEM career consideration.

Alongside these hypotheses, we explore the potential differences between general and subject-specific anxiety measures in relation to test performance and additional test-related variables (e.g., perceived test difficulty and

appeal). While these analyses are not focal in the present study, we believe they may nevertheless be informative and useful.

The results of the present study supplement the STEM education field, as studies regarding gender differences in the interplay between academic performance-related variables and STEM-related anxiety are scarce. In case the gender gap in science is related to anxiety, and other variables show correlations with performance and/or anxiety, the findings could inform academics and educational professionals about potential points for anxiety interventions, as outlined by Zakariya (2022).

Methodology

Sample and procedure

The target sample was 12th-grade students from general education schools and 3rd-year (partly 2nd-year) students from vocational educational institutions who had completed the national curriculum for the 4th level of education in physics, chemistry, biology, and geography. The goal was to include 20% (appx. $N=2500$) of the students from the respective general student population ($N=10147$). The data were collected in March 2023 from a representative sample of students with the aim of investigating science skills and science-related attitudes.

The survey was conducted in classrooms where each student solved the test and responded to the survey from an assigned computer. In total, the advised time for the procedure was 145 min, though this could be extended if necessary. 120 min were assigned solely for the science test administration, 15 min for the survey, and 10 min for a break between test sub-sections.

The students first took the science ability test. After that, the students filled out a three-part post-test survey. The first part focused on epistemic beliefs in science (not used in the present work), the second part was about science anxiety, and the third part regarded test-related attitudes. Importantly, in the part related to science anxiety, all students could respond to items regarding general science anxiety (please see the “Measures” section); subsequently, the students needed to select one subject to assess the related anxiety. Participating in the study was voluntary.

In total, 1907 Estonian 12th-graders took part in the study. After excluding the participants who had missing values, the sample comprised $N=1843$ students. Furthermore, students whose test-taking duration deviated more or less than three standard deviations from the average time were removed from the analysis. The effective sample comprised $N=1839$ students, 944 (51.3%) were female and 895 (48.7%) were male students. 1438 (78.2%) participants were from a general secondary school, while

401 (21.8%) were vocational secondary school students. 1466 (79.7%) students took the test in Estonian, and 373 (20.3%) students took the test in Russian language. Although the effective sample size was less than the planned $N=2500$, the sample was nevertheless representative of the student population of Estonia.

Measures

Science ability test The interdisciplinary science ability test consisted of physics, chemistry, biology, and geography assignments. These context-based multi-part tasks focused on science content and principles in daily life and global phenomena. The test was developed in several iterations across 2018 to 2022, based on both expert feedback as well as psychometric analysis; the detailed test development procedure is described in Vaino et al., (2024).

In total, the test consisted of 37 items which assessed the following facets:

- (a) Knowledge comprehension. Students were required to explain natural phenomena, identify cause-and-effect relationships, use scientific symbolism, perform calculations, and explain or construct scientific models;
- (b) Inquiry skills. Students demonstrated mastery of various research skills, from formulating research problems, questions, or hypotheses to evaluating the quality of conducted experiments;
- (c) Problem-solving. Students were tasked with resolving issues both with scientific content and those with a scientific basis and social relevance, making reasoned decisions;
- (d) Communication abilities. Students composed short texts on scientific topics, sought information from various sources, and assessed the reliability of the information obtained.

The responses were assessed based on the solution completion (e.g., 0=*incorrect*, 1=*partially correct*, 2=*correct*). Confirmatory factor analysis showed an acceptable model fit (Kline, 2015) for the unidimensional solution of the test, $\chi^2(629)=2268.29$, $p<0.001$, RMSEA=0.036, CFI=0.908, TLI=0.903 (Rannikmäe et al. 2023). In addition, Cronbach's $\alpha=0.849$ for the total test.

Post-testing survey general and subject-specific science anxiety. The science anxiety part of the procedure regarded the general and subject-specific statements. These items were created by the survey development team which included experts in education and psychology. The general science anxiety questionnaire included the following statements:

1. *Natural sciences lessons frighten me.*
2. *In natural science lessons, I worry that everything is too complicated for me.*
3. *When solving natural sciences homework, I worry that I do not understand it.*

The agreement with these statements was rated on a scale from 1 = *completely disagree* to 4 = *completely agree*. The scores were summed to form a general science anxiety score. Cronbach's alpha for this scale was $\alpha=0.86$.

In addition, the students could select one of the subjects (options: physics, chemistry, geography, biology) to assess their anxiety related to that subject. The following statements were used:

1. *I often worry that it is difficult for me in [chemistry/physics/geography/biology] lessons.*
2. *I get very tense when I have to learn [chemistry/physics/geography/biology] at home (do homeworks).*
3. *I get nervous when I am working on solving [chemistry/physics/geography/biology]-related problems.*
4. *I feel helpless when I am working on solving [chemistry/physics/geography/biology]-related problems.*
5. *I worry that I will get bad grades in [chemistry/physics/geography/biology].*

It is important to note that, as mentioned, only one subject could be chosen for evaluation. Hence, this scale was operationalized as "subject-specific" anxiety (as opposed to "general"). As with the general science anxiety inventory, the responses ranged from 1 = *completely disagree* to 4 = *completely agree*, and the scores of responses were summed. The internal reliability of the scale across the total sample was Cronbach's $\alpha=0.87$.

Post-testing survey science test-related attitudes. The students responded to seven items regarding test-taking motivation (1 = *completely disagree* to 5 = *completely agree*), which included statements regarding the importance of the test to the student (3 items; example item: "The test was important for me.") as well as the effort put into test-taking (4 items; example item: "I put in a lot of effort throughout the test"). Cronbach's $\alpha=0.70$ and $\alpha=0.84$ for test importance and effort, respectively. In addition, the students evaluated the difficulty (1 = *very simple* to 5 = *very difficult*) and appeal (1 = *very interesting* to 5 = *not at all interesting*) of the test with one item; the latter was reverse-coded for better interpretability to 1 = *not at all interesting* to 5 = *very interesting*).

Analysis

The data analysis was conducted in R v4.3.0 software (R Core Team, 2023). Descriptive statistics, internal reliability, and Pearson correlation analysis was conducted

with *psych* v2.3.3 (Revelle, 2021) and *RcmdrMisc* v.2.7–2 (Fox, 2022). Group differences in gender were analyzed with the independent samples t test from R’s *base()* package. The *lsr* v.0.5.2 (Navarro, 2015) was used to compute Cohen’s d-s for gender differences effect size estimation. We used common effect size benchmarks: $d=0.01$, 0.20, 0.50, 0.80, 1.20, and 2.00 for “very small”, “small”, “medium”, “large”, “very large”, and “huge” effect sizes, respectively (Cohen, 1988; Sawilowsky, 2009). However, we also want to point out that Kraft (2020) has argued that it may be more justified to interpret $d<0.05$, $d<0.20$, and $d>0.50$ as “small”, “medium”, and “large” effect sizes, respectively, in educational context. To test for differences between Pearson correlation coefficients for science test score and other variables, Fisher’s r-to-z transformation was employed, which converts correlations into normally distributed z-scores for comparison (Silver & Dunlap, 1987).

Results

In this section, we first present the descriptive statistics and correlation analysis results for the total sample. This is followed by the results on gender differences.

Descriptive statistics and correlations for the total sample

The descriptive statistics and correlations for the total sample are displayed in Table 1.

Table 1 shows that test performance is moderately and positively linked to test-taking duration: more time spent on solving the test is associated with better performance. Similarly, finding the test to be appealing and important

for oneself as well as reporting exerting more effort were linked to better performance. On the other hand, test difficulty had a small significant negative correlation with test performance. When it comes to test-taking duration, students who spent more time on test assignments also rated the importance of the test higher, found the test to be more appealing, and reported putting more effort into solving the assignments. These correlations were medium-to-large in effect sizes.

Looking into anxiety measures, the general science anxiety score did not correlate statistically significantly with other variables. Interestingly, general science anxiety did not correlate with subject-specific anxiety. On the other hand, subject-specific anxiety yields a small negative correlation with test performance, and students who rated the test more difficult also reported higher subject-specific anxiety. The results also show that students who found the test to be more difficult had more anxiety in the physics and chemistry domains.

When it comes to other test performance-related attitudes, the results in Table 1 show that finding the test appealing, in addition to positively correlating with better results and higher test-taking time, had a positive link with test importance and effort, and a negative association with test difficulty.

Gender differences in test performance-related factors and anxiety

Gender differences in science test performance and performance-related attitudes are displayed in Table 2 and

Table 1 Descriptive statistics and correlation coefficients

Variable	Descriptive statistics					Correlations						
	N	M	SD	Min	Max	1	2	3	4	5	6	7
1. Test score	1839	32.83	10.24	6	63	1						
2. Test duration (min)	1839	105.15	25.94	28	183	0.435***	1					
3. Test importance	1839	9.34	2.62	3	15	0.235***	0.325***	1				
4. Test effort	1839	13.65	3.53	4	20	0.286***	0.271***	0.618***	1			
5. Test difficulty	1839	3.70	0.79	1	5	-0.173***	-0.008	-0.036	-0.022	1		
6. Test appeal	1839	3.17	1.02	1	5	0.266***	0.250***	0.440***	0.412***	-0.166***	1	
7. GS anxiety	1839	7.21	2.66	3	12	-0.009	0.009	0.011	0.002	-0.035	0.018	1
8. SS anxiety ^a	1839	13.76	3.95	5	20	-0.125***	-0.049	-0.024	-0.056	0.186***	-0.078*	0.007
9. Physics anxiety	656	13.87	3.64	5	20	-0.119	-0.051	-0.077	-0.082	0.202***	-0.147**	0.052
10. Chemistry anxiety	795	14.65	3.85	5	20	-0.157***	-0.093	-0.025	-0.060	0.197***	-0.096	-0.002
11. Geography anxiety	139	11.65	4.04	5	20	0.025	0.089	0.012	-0.062	0.108	-0.090	-0.033
12. Biology anxiety	249	11.80	3.86	5	20	-0.122	0.014	0.009	-0.020	0.137	-0.003	-0.025

Correlation columns for 8–12 are not displayed because 9–12 are mutually exclusive and they form 8

^a Subject-specific science anxiety refers to the measure across the entire sample

GS general science anxiety, SS subject-specific science anxiety

P-values adjusted with the Holm’s method. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 2 Gender differences in test and anxiety-related variables

DV	Group	N	M	SD	t	df	p	d
1. Test score	Male	895	32.427	10.903	-1.620	1777.475	0.105	0.076
	Female	944	33.203	9.565				
2. Test duration	Male	895	101.126	26.662	-6.539	1805.723	<.001	0.306
	Female	944	108.967	24.646				
3. Test importance	Male	895	9.244	2.662	-1.578	1823.296	0.115	0.074
	Female	944	9.436	2.574				
4. Test effort	Male	895	13.642	3.599	-0.087	1822.347	0.931	0.004
	Female	944	13.657	3.470				
5. Test difficulty	Male	895	3.645	0.849	-2.837	1768.494	0.005	0.133
	Female	944	3.750	0.735				
6. Test appeal	Male	895	3.165	1.067	-0.263	1803.400	0.792	0.012
	Female	944	3.178	0.981				
7. GS anxiety	Male	895	7.082	2.627	-2.029	1834.932	0.043	0.095
	Female	944	7.333	2.679				
8. SS anxiety	Male	895	12.598	3.914	-12.782	1809.887	<.001	0.597
	Female	944	14.857	3.652				
9. Physics anxiety	Male	272	12.261	3.551	-10.106	551.891	<.001	0.813
	Female	384	15.010	3.529				
10. Chemistry anxiety	Male	408	13.716	3.880	-7.236	792.277	<.001	0.512
	Female	387	15.628	3.570				
11. Geography anxiety	Male	68	10.294	3.706	-4.086	136.959	<.001	0.692
	Female	71	12.944	3.938				
12. Biology anxiety	Male	147	11.184	3.734	-3.051	211.912	0.003	0.396
	Female	102	12.686	3.882				

DV dependent variable, GS general science anxiety, SS subject-specific science anxiety

Statistically significant differences ($p < .05$) are in **bold**

depicted in Fig. 1a, b. The figures display the results for the total sample, as well as for male and female students separately.

The results in Table 2 show that there were no gender differences in science test performance, nor was one group more motivated (i.e., assessing the test importance higher) than the other group to perform well on the test. In addition, girls and boys found the test similarly appealing, and both groups reported similar levels of effort in test performance.

The findings in Table 2, however, show that girls evaluated the test more difficult than boys. Female students reported higher anxiety levels in all anxiety measures.

How do different factors correlate with test performance across gender, and do these correlations differ? The latter question would allow further investigation if test-related variables or anxiety measures have a stronger link with test performance. To answer these questions, correlation coefficients were compared across the groups. The results are displayed in Table 3 and in Fig. 2a, b.

The results in Table 3 show that test duration had the strongest (positive) correlation with test performance,

i.e., more time spent on taking the test was correlated with better test results. As reported for the total sample in Table 1, test effort, importance, and appeal also had positive correlations with test performance also in both boys' and girls' samples. Test score and difficulty had a negative correlation. There were no gender differences in how strongly test-related variables correlated with test scores, indicating similar strengths in these associations across genders.

We also investigated the differences in correlations between test scores and various anxiety measures. Apart from general science anxiety (GS), geography and biology anxiety, other anxiety variables exhibited a statistically significant, albeit small, negative correlation with test scores in both boys and girls. However, the magnitudes of these correlations were similar, as the differences in correlations were not statistically significant.

To account for other possible effects (school type and test language), we conducted separate analyses on a homogeneous group of secondary school students who took the test in Estonian. The results, displayed in

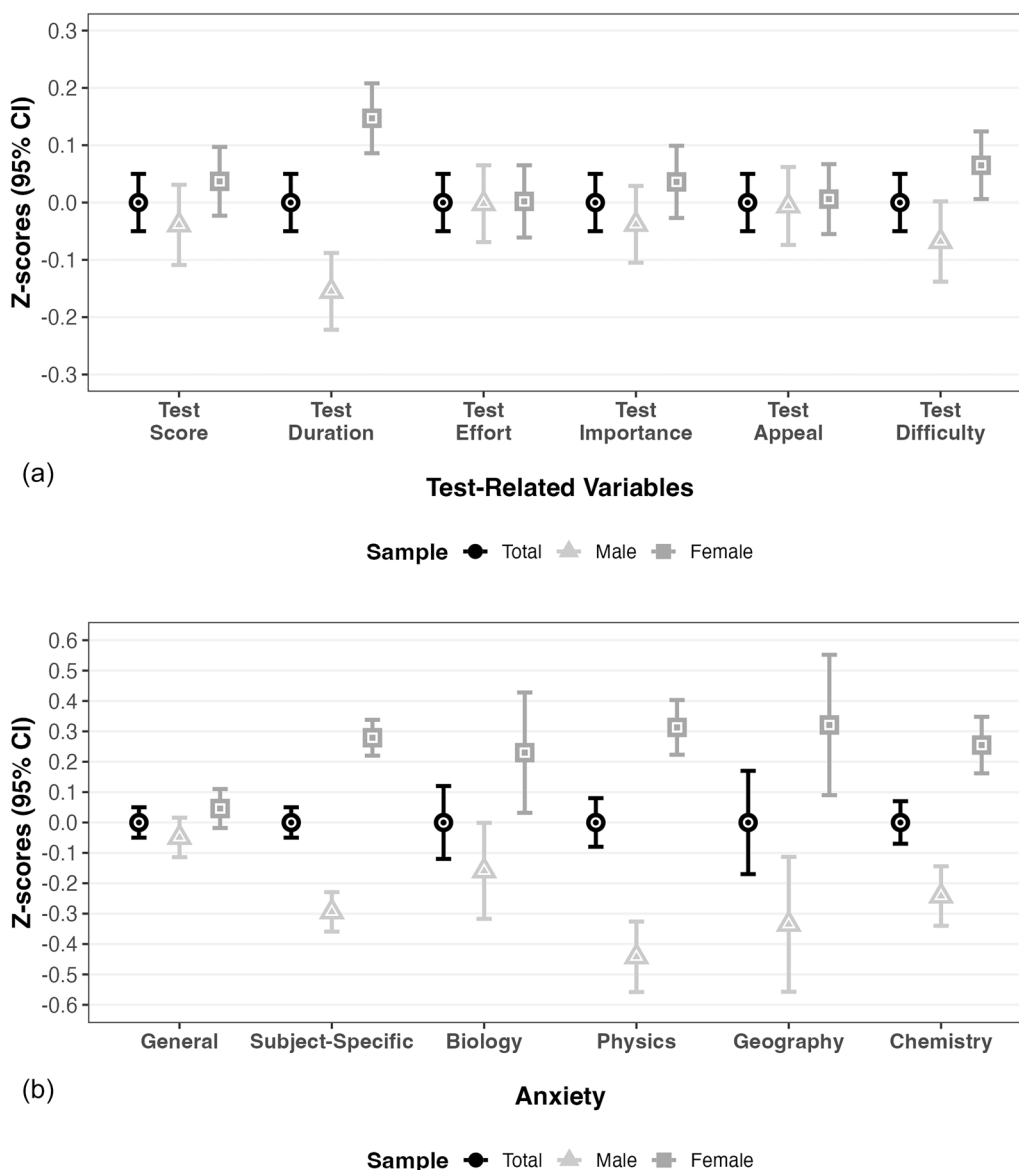


Fig. 1 **a** Error bar plots (95% CIs) for gender differences in science test performance, test-taking duration, and test-taking related attitudes. **b** Error bar plots (95% CIs) for group differences in general and subject-specific science anxiety

Supplementary Materials 1, showed marginal differences. Therefore, we discuss the total sample results as the main findings.

Discussion

The main aim of the present study is to explore the associations between science test performance, test-taking time, science anxiety, test-related attitudes, and gender differences. We posed several hypotheses, and conducted additional analyses.

According to the first hypothesis (H1), we expected that the science test scores are negatively associated with

science anxiety across the total sample. Previous works have shown that, among STEM subjects, mathematics anxiety generally predicts poorer mathematics achievement (Caviola et al., 2022; Guzmán et al., 2023; Namkung et al., 2019). The results showed that this hypothesis found partial support from the data. General science anxiety did not correlate with test performance; rather, the achievement-relevant anxiety seems to be subject-specific, especially when it comes to chemistry. Although the correlation was small, it nevertheless shows that students who reported higher chemistry-related anxiety also scored lower on a science test that includes elements of

Table 3 Correlation differences in links between test performance and other performance-related factors across different groups

DV	Group	N	r	p (r)	Z	p (diff)
1. Test duration	Male	895	0.454	<.001	0.807	0.420
	Female	944	0.411	<.001		
2. Test importance	Male	895	0.243	<.001	0.289	0.773
	Female	944	0.225	<.001		
3. Test effort	Male	895	0.316	<.001	1.023	0.306
	Female	944	0.254	<.001		
4. Test difficulty	Male	895	-0.157	<.001	0.673	0.501
	Female	944	-0.200	<.001		
5. Test appeal	Male	895	0.273	<.001	0.246	0.806
	Female	944	0.258	<.001		
6. GS anxiety	Male	895	-0.056	<.001	-1.417	0.156
	Female	944	0.038	0.001		
7. SS anxiety	Male	895	-0.173	<.001	-1.037	0.300
	Female	944	-0.105	0.001		
8. Physics anxiety	Male	272	-0.185	0.002	-0.426	0.670
	Female	384	-0.138	0.007		
9. Chemistry anxiety	Male	408	-0.197	<.001	-0.535	0.592
	Female	387	-0.145	0.004		
10. Geography anxiety	Male	68	0.018	0.885	0.150	0.881
	Female	71	-0.019	0.876		
11. Biology anxiety	Male	147	-0.154	0.063	-0.756	0.449
	Female	102	0.015	0.879		

DV dependent variable, GS general science anxiety, SS subject-specific science anxiety

Statistically significant correlations and correlation differences ($p < .05$) are in **bold**

other natural sciences. These findings also call for careful consideration of treating science anxiety as a general construct, whereas this anxiety might be subject-specific.

We expected that higher test importance, test-taking effort, time spent on test, and performance would all be positively correlated (H2 and H3). An explanation for such expectations lies in the assumption that students who perceive the science test as important to them, are more likely to put more effort into taking the test. Behaviorally, this could manifest in more time spent on solving the assignments, for example, making sure that the task is understood and double-checking the responses. In turn, this should result in higher test scores. Both hypotheses were supported by the data. In fact, test-taking time had the highest correlation with test performance among all the tested variables, yielding a medium-to-high effect. These results are also in line with literature that has previously demonstrated that more time taken for solving tasks tends to result in better performance (Silm et al., 2020). All other correlations were also positive, in the small-to-medium effect size range. Interestingly, girls took more time to solve the tasks than boys, but there were no differences in test importance and effort. That

said, the link between test score and test-taking duration did not differ across genders.

Finally, we expected not to see gender differences in test performance; however, girls were expected to score higher on science anxiety measures (H4). In addition, we expected that the correlation between science test performance and anxiety measures is stronger in girls than in boys (H5). The former hypothesis, too, found support from data. Although general science anxiety did not correlate with other measures, girls nevertheless scored higher, albeit with a small effect size. However, in subject-specific anxiety measures, the effect sizes of gender differences were medium-to-large. On the other hand, the latter hypothesis (H5) did not find support from the data, suggesting that the correlations between anxiety and performance do not differ statistically significantly between boys and girls. These results are interesting, as they seem to indicate that although girls report more anxiety, it does not seem to affect their results (or vice versa) more than those of boys. The findings also contradict those previously found in the mathematics domain (Devine et al., 2012; Dowker et al., 2016). Indeed, one could potentially hypothesize

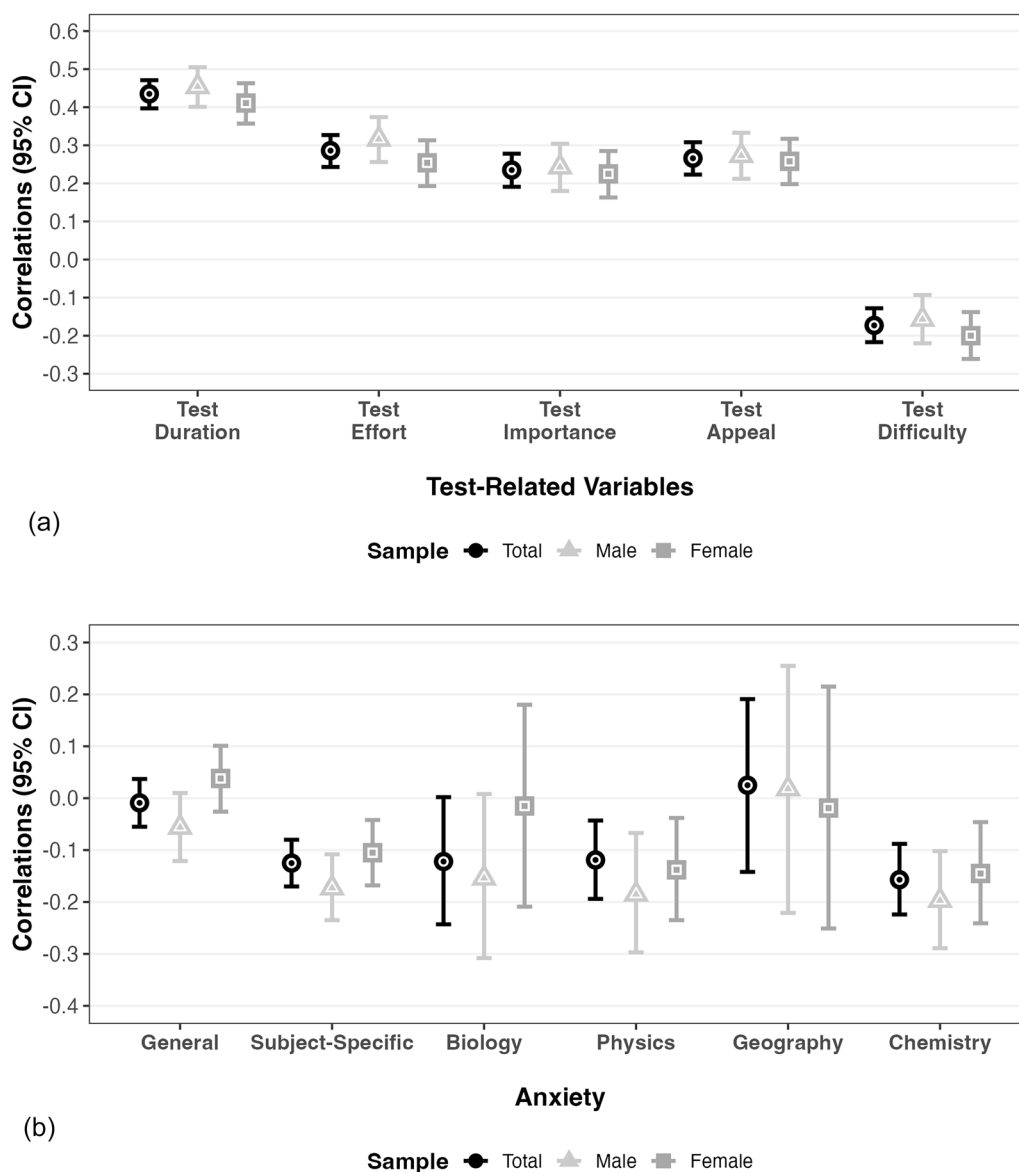


Fig. 2 **a** Error bar plots (95% CIs) for correlation differences between genders in science test performance, test-taking duration, and test-taking related attitudes. **b** Error bar plots (95% CIs) for correlation differences between genders in general and subject-specific science anxiety

that the gender gap in STEM careers (Schmader, 2023; Sevilla & Snodgrass Rangel, 2023)—at least based on science—might not originate from ability or from anxiety influencing performance, but rather from the anxiety itself.

There could be several explanations for these findings, ranging from societal expectations (e.g., stereotypes) to individual characteristics. Given that societal stereotypes affect not only how girls’ skills in STEM are perceived, but also how this perception negatively impacts their interest in these subjects (Master & Meltzoff, 2020; Master et al., 2021; Shapiro & Williams, 2012), as well as how

women may be perceived as lacking the qualities as scientists due to stereotypes (Carli et al., 2016), it might not be surprising that female students may feel the pressure to perform in STEM subjects while also having a lower STEM self-concept (Leibham et al., 2013). More recently, the role of social media has also been outlined as potentially influencing math and science attitudes (Daniels & Robnett, 2021).

Personality research has demonstrated that women, on average, score higher on the neuroticism personality trait, which reflects experiencing more negative affect, including anxious tendencies (Hofmann et al., 2023; Mac Giolla

& Kajonius, 2018). Relatedly, the neuroticism personality trait is not correlated with cognitive ability (Rozgonjuk et al., 2021a, 2021b). Hence, it may be that female students report more science anxiety than male students, despite their abilities not differing significantly. It has also been hypothesized that such anxious tendencies (at least in the STEM context) may drive a self-deprecating cycle in female students, where anxiety carries over to other testing situations (Pelch, 2018). This anxiety possibly affects performance and skews attitudes towards the study domain negatively. In this regard, it may be interesting to further examine the dynamics of experiencing anxiety and worrying tendencies in adolescents, and if there are general changes in anxious tendencies also in academic settings over time, similarly to more general personality characteristics (Möttus & Rozgonjuk, 2021).

With regard to additional exploratory analyses, we generally found that the correlations between test performance and other variables did not statistically significantly differ between boys and girls, meaning that the correlations found in the total sample should be applicable to both groups. The students who scored higher on the test also reported more positive feelings towards the test, i.e., found the test more appealing, and, conversely, found the test less difficult than the students who had a poorer test score. This is in line with previous findings, demonstrating that higher test-taking motivation, effort, and liking the test are associated with better achievement (Alhadabi & Karpinski, 2020; Pekrun et al., 2014; Silm et al., 2013; Živković et al., 2023), whereas the perceived difficulty of the test is linked to poorer results (Mazana et al., 2018).

Interestingly, those who found the test appealing, also took more time for solving the tasks. Although test difficulty and test appeal had a negative correlation, test difficulty did not predict test-taking time. These results suggest that students who took more time for solving the assignments did not seem to do so because they found the test very challenging; rather, also supported by the findings reported above, the findings indicate that the students were motivated to give their best. The findings may further underscore the importance of positive affective factors in successful test performance.

Additional interesting insights regard anxiety's negative correlation with test appeal and positive link with test difficulty. These links were subject-specific and the effects were rather small but they do seem to suggest that, in some cases, there might be an interplay between the perception of science assignments and anxiety. Granted that the present study did not investigate causal associations, one could further hypothesize that perceived test difficulty (or appeal) could be affected by the anxiety one has towards a science subject. Attitudes like finding

the subject difficult has been associated with anxiety in mathematics before (Rozgonjuk et al., 2020). Building positive attitudes towards science and testing can help boost science self-efficacy which, in turn, is associated with better results in science (Lau & Ho, 2022).

Science as a subject differs from mathematics, as it encompasses more domains. The results of our study showed, though, that when it comes to science attitudes (namely, anxiety), it may also be anchored to a specific subject, and not only rather than a general attitude towards science.

Studies tend to report that boys outperform girls in mathematics (Reilly et al., 2015; Rozgonjuk et al., 2023), but there is also evidence that girls outperform boys in science (OECD, 2019a). Regardless, female students tend to have poorer self-concept regarding these subjects, and could manifest in anxiety towards the subject (Balducci, 2023; Megreya et al., 2021). Our work is in line with the works regarding attitudes—girls do report higher science anxiety; on the other hand, our work also shows that this anxiety is not due to poorer achievement. Based on this, the findings suggest that STEM research might gain additional insights by also analyzing the different domains separately.

The main contribution of the present study was to provide empirical evidence for the 'S' in STEM research: science. Although science is an essential field in contemporary education and career opportunities, many STEM-focused papers seem to focus either on the 'M' in STEM—mathematics—or on STEM in general, including investigating gender differences. Our work revealed that there are not only interesting links and differences in the science domain—but science itself might be too general of a construct. Instead, specific science subjects could explain the link between performance, anxiety, and—perhaps also subsequent career considerations. The theoretical contribution is highlighting the subject-specific nature of science anxiety: it may be more nuanced. This knowledge can be used to refine theoretical models of (STEM-related) academic anxiety and performance. From a more practical perspective, knowing that girls have higher science anxiety but not necessarily lower ability could be used to form more positive attitudes in girls regarding science. For instance, although underestimated students may perform as well as others, they often have lower academic self-concept and expectations for success (Urhahne et al., 2011). This is also associated with poorer expectations from teachers (Urhahne et al., 2011). It could be argued that the societal expectations for gender roles (Hägglund & Leuze, 2021; Schmitt et al., 2008) may exacerbate the underestimation of the role of girls in science, potentially shaping the attitudes

of girls towards science and STEM in general. Hence, developing programs or approaches to inform the students and educators about girls being equally capable but more anxious in science than boys could be used towards building more realistic self-concepts regarding science. This may, in turn, help alleviate the gender gap in STEM.

The main limitation of this study is the cross-sectional design. Although anxiety toward science and specific subjects likely precedes test performance, the potential impact of test performance on other evaluations cannot be ruled out. Another limitation is relying on single items in some of the exploratory analyses. However, the items like test appeal and difficulty are relatively straightforward. In addition, studies have shown that single-item measures can yield results as reliable and valid as those of longer questionnaires (Diamantopoulos et al., 2012; Gardner et al., 1998; Rossiter, 2002). Even though test-taking duration might be an indication of test-taking effort (Silm et al., 2020), we acknowledge that we cannot definitively distinguish between rapid guessing, proficiency, diligence, and difficulty in answering. Although out of the scope in the present study, such distinction could be made by investigating the link between response patterns and timestamps of response events. While the anxiety-related items used in this study mostly focused on how the students generally felt about typical science-related educational situations (e.g., regarding homeworks), it should also be noted that the potential test anxiety experienced by students could also have influenced the post-test science anxiety evaluations. Future studies could consider implementing both before and after assessments.

The role of technology in STEM education can also be investigated in light of the present results. For instance, the use of virtual reality (Yang et al., 2024), social media (Achilleos et al., 2019; He et al., 2016), as well as smartphones (Mella-Norambuena et al., 2021; Smith et al., 2023) has shown promise in STEM education with regard to improving students' motivation. More research is needed to investigate gender differences in technology use in STEM settings. In addition, the use of digital technology in classroom settings should be done with caution (Aru & Rozgonjuk, 2022; Rozgonjuk, Täht, et al., 2021), as studies have shown that social media and smartphone use, as well as notifications received from these media and devices, are linked to more procrastination (Rozgonjuk et al., 2018a, 2018b), superficial learning styles (Rozgonjuk et al., 2019; Rozgonjuk, Saal, et al., 2018), and boredom proneness (Elhai et al., 2021; Wolniewicz et al., 2020). Hence, non-skillful implementation of technology assistance may not necessarily improve the attitudes towards science.

Finally, it would be interesting to compare the science-related variables of this study with other STEM components, such as engineering and mathematics, as well as in interdisciplinary contexts (Darmawansah et al., 2023; Gao et al., 2020).

Conclusions

In this study, we focused on the interplay and gender differences in the S-domain of STEM: science. Although some of the findings were in line with previous results from the general STEM and mathematics education fields, there are also indications that "science" might not be a broad concept. Anxiety toward science might be subject-specific. Furthermore, the results showed that although there were no gender differences in terms of science test performance, girls reported consistently higher levels of anxiety, both in terms of general and subject-specific science. However, the correlations between anxiety measures and performance were not stronger in girls than in boys. The results suggest that the gender gap in STEM might not necessarily stem from ability but rather motivational variables, such as anxiety. However, additional research is needed to establish the source of science-related anxiety in girls and its potential impact on pursuing a STEM career.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40594-024-00504-4>.

Additional file 1.

Acknowledgements

The authors would like to thank the Education and Youth Board of Estonia for providing the data. Dr. Rozgonjuk was affiliated with University of Tartu at the time of conducting this study. He now works in the private sector, and his contribution to this paper was made independently during his free time.

Author contributions

DR: study conceptualization, data analysis, writing the first manuscript draft; KT: study conceptualization, critical review; RS: critical review; MT: critical review; MR: study conceptualization, critical review.

Funding

This study was supported by the European Union's Horizon 2020 project „Sci-Car” (Addressing Attractiveness of Science Career Awareness) No. 952470. Dr. Täht's work was supported by the Estonian Research Council Grant PRG2190.

Availability of data and materials

The aggregated data used in the present study will be considered to be made available upon scholarly request.

Declarations

Ethics approval and consent to participate

The Education and Youth Board of Estonia ordered the study as part of the national educational policy assessment. The procedure of the study was in accordance with national legislation and ethical principles. That includes the informed consent regarding voluntary participation in the study from the participants.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Received: 3 March 2024 Accepted: 26 August 2024

Published online: 17 September 2024

References

- Achilleos, A. P., Mettouris, C., Yeratziotis, A., Papadopoulos, G. A., Pllana, S., Huber, F., Jager, B., Leitner, P., Ocsosvsky, Z., & Dinnyes, A. (2019). SciChallenge: A social media aware platform for contest-based STEM education and motivation of young students. *IEEE Transactions on Learning Technologies*, 12(1), 98–111. <https://doi.org/10.1109/TLT.2018.2810879>
- Alhadabi, A., & Karpinski, A. C. (2020). Grit, self-efficacy, achievement orientation goals, and academic performance in University students. *International Journal of Adolescence and Youth*, 25(1), 519–535. <https://doi.org/10.1080/02673843.2019.1679202>
- Andersen, L. (2014). Visual-spatial ability: important in STEM. *Ignored in Gifted Education. Roeper Review*, 36(2), 114–121. <https://doi.org/10.1080/02783193.2014.884198>
- Anni, K., Vainik, U., & Möttus, R. (2023). Personality profiles of 263 occupations. *PsyArXiv*. <https://doi.org/10.31234/osf.io/ajvg2>
- Aru, J., & Rozgonjuk, D. (2022). The effect of smartphone use on mental effort, learning, and creativity. *Trends in Cognitive Sciences*, 26(10), 821–823. <https://doi.org/10.1016/j.tics.2022.07.002>
- Balducci, M. (2023). Linking gender differences with gender equality: A systematic-narrative literature review of basic skills and personality. *Frontiers in Psychology*, 14, 1105234. <https://doi.org/10.3389/fpsyg.2023.1105234>
- Baloglu, M., & Kocak, R. (2006). A multivariate investigation of the differences in mathematics anxiety. *Personality and Individual Differences*. <https://doi.org/10.1016/j.paid.2005.10.009>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H.: Freeman.
- Berkowitz, M., & Stern, E. (2018). Which cognitive abilities make the difference? predicting academic achievements in advanced STEM studies. *Journal of Intelligence*, 6(4), 48. <https://doi.org/10.3390/jintelligence6040048>
- Burns, E. C., Martin, A. J., Kennett, R. K., Pearson, J., & Munro-Smith, V. (2021). Optimizing science self-efficacy: A multilevel examination of the moderating effects of anxiety on the relationship between self-efficacy and achievement in science. *Contemporary Educational Psychology*, 64, 101937. <https://doi.org/10.1016/j.cedpsych.2020.101937>
- Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review*, 106(4), 676.
- Carli, L. L., Alawa, L., Lee, Y., Zhao, B., & Kim, E. (2016). Stereotypes about gender and science: Women ≠ scientists. *Psychology of Women Quarterly*, 40(2), 244–260. <https://doi.org/10.1177/0361684315622645>
- Caviola, S., Toffalini, E., Giofrè, D., Ruiz, J. M., Szűcs, D., & Mammarella, I. C. (2022). Math performance and academic anxiety forms, from sociodemographic to cognitive aspects: A Meta-analysis on 906,311 participants. *Educational Psychology Review*, 34(1), 363–399. <https://doi.org/10.1007/s10648-021-09618-5>
- Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, 143(1), 1–35. <https://doi.org/10.1037/bul0000052>
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences. L. Erlbaum Associates. <https://doi.org/10.4324/9780203771587>
- Cotner, S., Jenö, L. M., Walker, J. D., Jørgensen, C., & Vandvik, V. (2020). Gender gaps in the performance of Norwegian biology students: The roles of test anxiety and science confidence. *International Journal of STEM Education*, 7(1), 55. <https://doi.org/10.1186/s40594-020-00252-1>
- Daniels, E. A., & Robnett, R. D. (2021). The STEM pipeline: Do media and objectified body consciousness create an early exit for middle school girls? *The Journal of Early Adolescence*, 41(7), 1099–1124. <https://doi.org/10.1177/0272431620983442>
- Darmawansah, D., Hwang, G.-J., Chen, M.-R.A., & Liang, J.-C. (2023). Trends and research foci of robotics-based STEM education: A systematic review from diverse angles based on the technology-based learning model. *International Journal of STEM Education*, 10(1), 12. <https://doi.org/10.1186/s40594-023-00400-3>
- Devine, A., Fawcett, K., Szűcs, 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), 1–9.
- Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: A predictive validity perspective. *Journal of the Academy of Marketing Science*, 40(3), 434–449. <https://doi.org/10.1007/s11747-011-0300-3>
- Dökme, İ, Açıksöz, A., & Koyunlu Ünlü, Z. (2022). Investigation of STEM fields motivation among female students in science education colleges. *International Journal of STEM Education*, 9(1), 8. <https://doi.org/10.1186/s40594-022-00326-2>
- Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2016.00508>
- Early, E., Miller, S., Dunne, L., Thurston, A., & Filiz, M. (2020). The influence of socio-economic background and gender on school attainment in the United Kingdom: A systematic review. *Review of Education*, 8(1), 120–152. <https://doi.org/10.1002/rev3.3175>
- Elhai, J. D., Rozgonjuk, D., Alghraibeh, A. M., & Yang, H. (2021). Disrupted daily activities from interruptive smartphone notifications: Relations with depression and anxiety severity and the mediating role of boredom proneness. *Social Science Computer Review*. <https://doi.org/10.1177/0894439319858008>
- Foley, A. E., Herts, J. B., Boronovi, F., Guerriero, S., Levine, S. C., & Beilock, S. L. (2017). The math anxiety-performance link: A global phenomenon. *Current Directions in Psychological Science*, 26(1), 52–58. <https://doi.org/10.1177/0963721416672463>
- Fox, J. (2022). *RcmdrMisc: R Commander Miscellaneous Functions* (Version 2.7–2). <https://CRAN.R-project.org/package=RcmdrMisc>
- Gao, X., Li, P., Shen, J., & Sun, H. (2020). Reviewing assessment of student learning in interdisciplinary STEM education. *International Journal of STEM Education*, 7(1), 24. <https://doi.org/10.1186/s40594-020-00225-4>
- Gardner, D. G., Cummings, L. L., Dunham, R. B., & Pierce, J. L. (1998). Single-Item versus multiple-item measurement scales: An empirical comparison. *Educational and Psychological Measurement*, 58(6), 898–915. <https://doi.org/10.1177/0013164498058006003>
- Glynn, S. M., Taasobshirazi, G., & Brickman, P. (2009). Science Motivation Questionnaire: Construct validation with nonscience majors. *Journal of Research in Science Teaching*, 46(2), 127–146. <https://doi.org/10.1002/tea.20267>
- Guzmán, B., Rodríguez, C., & Ferreira, R. A. (2023). Effect of parents' mathematics anxiety and home numeracy activities on young children's math performance-anxiety relationship. *Contemporary Educational Psychology*, 72, 102140. <https://doi.org/10.1016/j.cedpsych.2022.102140>
- Hägglund, A. E., & Leuze, K. (2021). Gender differences in STEM expectations across countries: How perceived labor market structures shape adolescents' preferences. *Journal of Youth Studies*, 24(5), 634–654. <https://doi.org/10.1080/13676261.2020.1755029>
- He, L., Murphy, L., & Luo, J. (2016). Using social media to promote STEM Education: Matching college students with role models. In B. Berendt, B. Bringmann, É. Fromont, G. Garriga, P. Miettinen, N. Tatti, & V. Tresp (Eds.), *Machine Learning and Knowledge Discovery in Databases* (pp. 79–95). United States: Springer International Publishing.
- Hofmann, R., Rozgonjuk, D., Soto, C. J., Ostendorf, F., & Möttus, R. (2023). There are a million ways to be a woman and a million ways to be a man: Gender differences across personality nuances and nations. *PsyArXiv*. <https://doi.org/10.31234/osf.io/cedwk>
- Kayan-Fadllemula, F., Sellami, A., Abdelkader, N., & Umer, S. (2022). A systematic review of STEM education research in the GCC countries: Trends, gaps and barriers. *International Journal of STEM Education*, 9(1), 2. <https://doi.org/10.1186/s40594-021-00319-7>
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. New York: Guilford publications.
- Lau, K.-C., & Ho, S.-C.E. (2022). Attitudes towards Science, teaching practices, and science performance in PISA 2015: Multilevel analysis of the Chinese and western top performers. *Research in Science Education*, 52(2), 415–426. <https://doi.org/10.1007/s11165-020-09954-6>
- Leibham, M. B., Alexander, J. M., & Johnson, K. E. (2013). Science interests in preschool boys and girls: Relations to later self-concept and science

- achievement. *Science Education*, 97(4), 574–593. <https://doi.org/10.1002/sce.21066>
- Mac Giolla, E., & Kajonius, P. J. (2018). Sex differences in personality are larger in gender equal countries: Replicating and extending a surprising finding. *International Journal of Psychology*. <https://doi.org/10.1002/ijop.12529>
- Li, Y., Wang, K., Xiao, Y., & Froyd, J. E. (2020). Research and trends in STEM education: A systematic review of journal publications. *International Journal of STEM Education*, 7(1), 11. <https://doi.org/10.1186/s40594-020-00207-6>
- Mallow J.V. (2006) Science anxiety: Research and action. *Handbook of college science teaching*, 3–14
- Master, A. H., & Meltzoff, A. N. (2020). Cultural stereotypes and sense of belonging contribute to gender gaps in STEM. *International Journal of Gender, Science and Technology*, 12(1), 152–198.
- Master, A. H., Meltzoff, A. N., & Cheryan, S. (2021). Gender stereotypes about interests start early and cause gender disparities in computer science and engineering. *Proceedings of the National Academy of Sciences*, 118(48), e2100030118. <https://doi.org/10.1073/pnas.2100030118>
- Mazana, M. Y., Montero, C. S., & Casimir, R. O. (2018). Investigating students' attitude towards learning mathematics. *International Electronic Journal of Mathematics Education*. <https://doi.org/10.29333/iejme/3997>
- McKinney, J., Chang, M.-L., & Glassmeyer, D. (2021). Why females choose STEM majors: Understanding the relationships between major, personality, interests, self-efficacy, and anxiety. *Journal for STEM Education Research*, 4(3), 278–300. <https://doi.org/10.1007/s41979-021-00050-6>
- Megreya, A. M., Szűcs, D., & Moustafa, A. A. (2021). The abbreviated science anxiety scale: Psychometric properties, gender differences and associations with test anxiety, general anxiety and science achievement. *PLoS ONE*, 16(2), e0245200. <https://doi.org/10.1371/journal.pone.0245200>
- Mella-Norambuena, J., Cobo-Rendon, R., Lobos, K., Sáez-Delgado, F., & Maldonado-Trapp, A. (2021). Smartphone use among undergraduate STEM students during COVID-19: An opportunity for higher education? *Education Sciences*, 11(8), 417. <https://doi.org/10.3390/educsci11080417>
- Möttus, R., & Rozgonjuk, D. (2021). Development is in the details: Age differences in the Big Five domains, facets, and nuances. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/pspp0000276>
- 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>
- Navarro D. (2015) *Learning statistics with R: A tutorial for psychology students and other beginners*. <http://health.adelaide.edu.au/psychology/ccs/teaching/Isr/>.
- OECD. (2019a). Chapter 7 girls' and boys' performance in PISA In PISA 2018 results (Volume II): where all students can succeed. OECD. <https://doi.org/10.1787/b5fd1b8f-en>
- OECD. (2019b). TALIS 2018 results (volume I): Teachers and school leaders as lifelong learners. OECD. <https://doi.org/10.1787/1d0bc92a-en>
- OECD. (2024) *OECD Dashboard on Gender Gaps*. <https://oecd.org/stories/gender/dashboard>
- Özcan, Z. Ç., & Eren Gümüş, A. (2019). A modeling study to explain mathematical problem-solving performance through metacognition, self-efficacy, motivation, and anxiety. *Australian Journal of Education*, 63(1), 116–134. <https://doi.org/10.1177/0004944119840073>
- Pekrun, R., Hall, N. C., Goetz, T., & Perry, R. P. (2014). Boredom and academic achievement: Testing a model of reciprocal causation. *Journal of Educational Psychology*, 106(3), 696–710. <https://doi.org/10.1037/a0036006>
- Pelch, M. (2018). Gendered differences in academic emotions and their implications for student success in STEM. *International Journal of STEM Education*. <https://doi.org/10.1186/s40594-018-0130-7>
- Potvin, P., & Hasni, A. (2014). Interest, motivation and attitude towards science and technology at K-12 levels: A systematic review of 12 years of educational research. *Studies in Science Education*, 50(1), 85–129. <https://doi.org/10.1080/03057267.2014.881626>
- Potvin, P., Hasni, A., Sy, O., & Riopel, M. (2020). Two crucial years of science and technology schooling: A longitudinal study of the major influences on and interactions between self-concept, interest, and the intention to pursue S&T. *Research in Science Education*, 50(5), 1739–1761. <https://doi.org/10.1007/s11165-018-9751-6>
- Rannikmäe, M., Vaino, K., Soobard, R., Teppo, M., & Reisenbuk, E. (2023). Lühikokkuvõte 2022/2023. Õppeaasta loodusainete IV kooliastme katselise tasemetöö tulemustest [A brief summary of the proficiency test results for natural sciences in the IV grade level for the 2022/2023 academic year]. Education and Youth Board of Estonia.
- R Core Team. (2023) *R: A language and environment for statistical computing* (Version 4.3.0) [Computer software]. R Core Team.
- Reilly, D., Neumann, D. L., & Andrews, G. (2015). Sex differences in mathematics and science achievement: A meta-analysis of national assessment of educational progress assessments. *Journal of Educational Psychology*, 107(3), 645–662. <https://doi.org/10.1037/edu0000012>
- Revelle W. (2021). *psych: Procedures for personality and psychological research*. (Version 2.2.3) [Computer software]. <https://CRAN.R-project.org/package=psych>
- Rios, J. A., Deng, J., & Ihlenfeldt, S. D. (2022). To what degree does rapid guessing distort aggregated test scores? *A Meta-Analytic Investigation*. *Educational Assessment*, 27(4), 356–373. <https://doi.org/10.1080/10627197.2022.2110465>
- Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, 19(4), 305–335. [https://doi.org/10.1016/S0167-8116\(02\)00097-6](https://doi.org/10.1016/S0167-8116(02)00097-6)
- Rozgonjuk, D., Elhai, J. D., Ryan, T., & Scott, G. G. (2019). Fear of missing out is associated with disrupted activities from receiving smartphone notifications and surface learning in college students. *Computers Education*, 140, 103590. <https://doi.org/10.1016/j.compedu.2019.05.016>
- Rozgonjuk, D., Kattago, M., & Täht, K. (2018a). Social media use in lectures mediates the relationship between procrastination and problematic smartphone use. *Computers in Human Behavior*, 89, 191–198. <https://doi.org/10.1016/j.chb.2018.08.003>
- Rozgonjuk, D., Saal, K., & Täht, K. (2018b). Problematic smartphone use, deep and surface approaches to learning, and social media use in lectures. *International Journal of Environmental Research and Public Health*. <https://doi.org/10.3390/ijerph15010092>
- Rozgonjuk, D., Konstabel, K., Barker, K., Rannikmäe, M., & Täht, K. (2023). Epistemic beliefs in science, socio-economic status, and mathematics and science test results in lower secondary education: A multilevel perspective. *Educational Psychology*, 43(1), 22–37. <https://doi.org/10.1080/01443410.2022.2144143>
- Rozgonjuk, D., Kraav, T., Mikkor, K., Orav-Puurand, K., & Täht, K. (2020). Mathematics anxiety among STEM and social sciences students: The roles of mathematics self-efficacy, and deep and surface approach to learning. *International Journal of STEM Education*. <https://doi.org/10.1186/s40594-020-00246-z>
- Rozgonjuk, D., Schmitz, F., Kannen, C., & Montag, C. (2021a). Cognitive ability and personality: Testing broad to nuanced associations with a smartphone app. *Intelligence*, 88, 101578. <https://doi.org/10.1016/j.intell.2021.101578>
- Rozgonjuk, D., Täht, K., & Vassil, K. (2021b). Internet use at and outside of school in relation to low- and high-stakes mathematics test scores across 3 years. *International Journal of STEM Education*. <https://doi.org/10.1186/s40594-021-00287-y>
- Sakellariou, C., & Fang, Z. (2021). Self-efficacy and interest in STEM subjects as predictors of the STEM gender gap in the US: The role of unobserved heterogeneity. *International Journal of Educational Research*, 109, 101821. <https://doi.org/10.1016/j.ijer.2021.101821>
- Sawilowsky, S. S. (2009). New Effect size rules of Thumb. *Journal of Modern Applied Statistical Methods*. <https://doi.org/10.22237/jmasm/1257035100>
- Schmader, T. (2023). Gender inclusion and fit in STEM. *Annual Review of Psychology*, 74(1), 219–243. <https://doi.org/10.1146/annurev-psych-032720-043052>
- Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman? sex differences in Big Five personality traits across 55 cultures. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/0022-3514.94.1.168>
- Sevilla, M. P., & Snodgrass Rangel, V. (2023). Gender differences in STEM career development in postsecondary vocational-technical education. a social cognitive career theory test. *Journal of Career Development*, 50(2), 255–272. <https://doi.org/10.1177/08948453221086979>
- Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex Roles*, 66(3–4), 175–183. <https://doi.org/10.1007/s11199-011-0051-0>
- Silm, G., Must, O., & Täht, K. (2013). Test-taking effort as a predictor of performance in low-stakes tests. *Trames Journal of the Humanities and Social Sciences*. <https://doi.org/10.3176/tr.2013.4.08>

- Silm, G., Pedaste, M., & Täht, K. (2020). The relationship between performance and test-taking effort when measured with self-report or time-based instruments: A meta-analytic review. *Educational Research Review*, 31, 100335. <https://doi.org/10.1016/j.edurev.2020.100335>
- Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fisher's z transformation be used? *Journal of Applied Psychology*, 72(1), 146–148. <https://doi.org/10.1037/0021-9010.72.1.146>
- Smith, J. R., Snapp, B., Madar, S., Brown, J. R., Fowler, J., Andersen, M., Porter, C. D., & Orban, C. (2023). A smartphone-based virtual reality plotting system for STEM education. *Primus*, 33(1), 1–15. <https://doi.org/10.1080/10511970.2021.2006378>
- So, W. W. M., Chen, Y., & Chow, S. C. F. (2022). Primary school students' interests in STEM careers: How conceptions of STEM professionals and gender moderation influence. *International Journal of Technology and Design Education*, 32(1), 33–53. <https://doi.org/10.1007/s10798-020-09599-6>
- Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science*, 29(4), 581–593. <https://doi.org/10.1177/0956797617741719>
- Täht, K., Mikkor, K., Aaviste, G., & Rozgonjuk, D. (2023). What motivates and demotivates Estonian mathematics teachers to continue teaching? The roles of self-efficacy, work satisfaction, and work experience. *Journal of Mathematics Teacher Education*. <https://doi.org/10.1007/s10857-023-09587-2>
- Tandrayen-Ragoobur, V., & Gokulsing, D. (2022). Gender gap in STEM education and career choices: What matters? *Journal of Applied Research in Higher Education*, 14(3), 1021–1040. <https://doi.org/10.1108/JARHE-09-2019-0235>
- Tobias, S., & Weissbrod, C. (1980). Anxiety and mathematics: An update. *Harvard Educational Review*. <https://doi.org/10.17763/haer.50.1.xw483257j6035084>
- Treialt SL. (2021). *Women in STEM fields, career choices and gender pay gap: Evidence from the Baltics*.
- Urhahne, D., Chao, S.-H., Florineth, M. L., Luttenberger, S., & Paechter, M. (2011). Academic self-concept, learning motivation, and test anxiety of the underestimated student: The underestimated student. *British Journal of Educational Psychology*, 81(1), 161–177. <https://doi.org/10.1348/000709910X504500>
- Vaino, K., Rosin, T., Liiber, Ü., Soobard, R., Teppo, M., Valdmann, A., Reisenbuk, E., & Rannikmäe, M. (2024). The development of a national e-test on science competence for the third school level: An assessment to support learning. *Eesti Haridusteaduste Ajakiri Estonian Journal of Education*, 12(1), 88–120. <https://doi.org/10.12697/eha.2024.12.1.05>
- Verdugo-Castro, S., García-Holgado, A., & Sánchez-Gómez, M. C. (2022). The gender gap in higher STEM studies: A systematic literature review. *Helion*, 8(8), e10300. <https://doi.org/10.1016/j.helion.2022.e10300>
- Vos, H., Marinova, M., De Léon, S. C., Sasanguie, D., & Reynvoet, B. (2023). Gender differences in young adults' mathematical performance: Examining the contribution of working memory, math anxiety and gender-related stereotypes. *Learning and Individual Differences*, 102, 102255. <https://doi.org/10.1016/j.lindif.2022.102255>
- Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*. <https://doi.org/10.1037/a0036620>
- Wang, N., Tan, A.-L., Zhou, X., Liu, K., Zeng, F., & Xiang, J. (2023). Gender differences in high school students' interest in STEM careers: A multi-group comparison based on structural equation model. *International Journal of STEM Education*, 10(1), 59. <https://doi.org/10.1186/s40594-023-00443-6>
- Warren, L., Reilly, D., Herdan, A., & Lin, Y. (2020). Self-efficacy, performance and the role of blended learning. *Journal of Applied Research in Higher Education*, 13(1), 98–111. <https://doi.org/10.1108/JARHE-08-2019-0210>
- Widya, R., & R., & Laila Rahmi, Y. (2019). STEM education to fulfil the 21st century demand: A literature review. *Journal of Physics: Conference Series*, 1317(1), 012208. <https://doi.org/10.1088/1742-6596/1317/1/012208>
- Wolniewicz, C. A., Rozgonjuk, D., & Elhai, J. D. (2020). Boredom proneness and fear of missing out mediate relations between depression and anxiety with problematic smartphone use. *Human Behavior and Emerging Technologies*, 2, 61–70. <https://doi.org/10.1002/hbe2.159>
- Yang, C., Zhang, J., Hu, Y., Yang, X., Chen, M., Shan, M., & Li, L. (2024). The impact of virtual reality on practical skills for students in science and engineering education: A meta-analysis. *International Journal of STEM Education*, 11(1), 28. <https://doi.org/10.1186/s40594-024-00487-2>
- Zakariya, Y. F. (2022). Improving students' mathematics self-efficacy: A systematic review of intervention studies. *Frontiers in Psychology*, 13, 986622. <https://doi.org/10.3389/fpsyg.2022.986622>
- 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, 552355. <https://doi.org/10.3389/fpsyg.2020.552355>
- Živković, M., Pellizzoni, S., Doz, E., Cuder, A., Mammarella, I., & Passolunghi, M. C. (2023). Math self-efficacy or anxiety? The role of emotional and motivational contribution in math performance. *Social Psychology of Education*, 26(3), 579–601. <https://doi.org/10.1007/s11218-023-09760-8>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.