



Gender stereotypes: implicit threat to performance or boost for motivational aspects in primary school?

Johanna Maria Hermann¹ · Regina Vollmeyer¹

Received: 16 August 2021 / Accepted: 5 April 2022 / Published online: 28 April 2022
© The Author(s) 2022

Abstract

Based on stereotype threat and stereotype lift theory, this study explores implicit stereotype threat effects of gender stereotypes on the performance of primary school children in mathematics. Moreover, effects of implicit gender stereotypical cues (gender-specific task material) on motivational aspects were explored, which have revealed mixed results in stereotype threat research in the past. $N=151$ German primary school children (47.7% female; mean age: $M=9.81$, $SD=0.60$) calculated either stereotypical or neutral mathematical text problems before motivational aspects were assessed. Contradicting our expectations, results neither revealed a stereotype threat effect on girls' performance nor a lift effect on the boys. Instead, girls calculating stereotypical tasks outperformed girls in the control group, whereas boys' performance did not significantly differ compared to the control group. Regarding motivational aspects, only traditional gender differences emerged as girls reported significantly more pressure and tension calculating the mathematical tasks. The discussion focuses on the way in which stereotypes can affect children's cognitive performance and in turn, their mathematical performance.

Keywords Implicit · Stereotype threat · Gender-specific task material · Mathematics · Motivation

✉ Johanna Maria Hermann
Hermann@psych.uni-frankfurt.de

Regina Vollmeyer
R.Vollmeyer@paed.psych.uni-frankfurt.de

¹ Department of Psychology, Educational Psychology, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 6, 60629 Frankfurt am Main, Germany

1 Introduction

Pink doll dresses and blue toy blocks—children are surrounded by gender¹ specific, often gender stereotypical material every day of their life (MacPhee & Prendergast, 2019; Murnen et al., 2016). Whilst a lot of research has been done on children’s preference for gender-stereotypical toys (e.g. Spinner et al., 2018), very little is known so far about the effects of gender-stereotypical task materials on academic performance. However, there has been a growing trend in the last decades—especially in the United States—to argue for gender-specific education (for an overview see Bigler & Signorella, 2011). Corresponding policies reach from single training programs for prospective teachers in gender-specific instruction via individual gender-segregated classes to ongoing gender-segregated schooling in all subjects (Pahlke & Hyde, 2016). Some proponents of these ideas argue that sexism in coeducation hinders girls and boys from performing up to their full potential (e.g. Salomone, 2004). Other advocates state substantial biological gender differences between boys and girls, resulting in gender-specific learning styles (e.g. Gurian et al., 2001; Sax, 2006).

Similar approaches to gender-specific education have been discussed in Germany (Faulstich-Wieland, 2011), predominantly with a focus on school subjects, showing gender-related differences resistant to change in large-scale assessment. As a result, gender-segregated classes for girls were most often the subject of research with regard to mathematics (e.g. Rudolph-Albert & Keller, 2007) and physics (e.g. Hannover & Kessels, 2002; Häussler & Hoffmann, 2002). Meanwhile, separate language classes are also suggested for boys (Budde et al., 2016), as they are considered as “left behind” in the educational system (Hannover & Kessels, 2011, p.89), consistent with the “boys’ crisis” in the United States (e.g. Kleinfeld, 2009, p.113). Supporters of these positions in Germany argue that, on the one hand, instruction in the natural sciences is designed to appeal only to boys (Budde et al., 2016), while on the other hand the “feminisation” of the whole school context (Heyder & Kessels, 2013, p.605) is discussed, discriminating against boys systematically (Guggenbühl, 2008). However, in contrast to the United States, in Germany most gender-specific programs are offered outside of the school system (Budde et al., 2016). In this spirit, gender-specific task materials for home learning environments were also developed, aimed at a growing market of popular science literature on educational questions. Such educational books have stereotypical titles such as ‘100 mathematical tasks that really engage girls’ interest’ (Speicher, 2009a) or, respectively, offer essay exercises and dictations specifically designed for boys. Whereas girls are asked to calculate how many tickets are left for a ballet show, boys have to find out the number of remaining tickets for a soccer match (Speicher, 2009b). Empirical research concerning the success of gender-specific educational policies revealed mixed results in

¹ We focus only on binary gender perceptions as we are interested in the influence of gender stereotypes associated with the female and male gender. However, we acknowledge, that the conception of gender as binary is a narrow conception that not necessarily reflects the full range of possible gender identifications.

Germany (Faulstich-Wieland, 2011) and only small effect sizes in an international meta-analysis (Pahlke et al., 2014), whereas gender-specific materials are seldom systematically explored. Furthermore, some critics even suspect gender-segregated education to foster stereotypical thinking in children and adolescents instead of promoting academic interest or a performance change between girls and boys (Datnow et al., 2001; Fabes et al., 2013, 2015; Hilliard & Liben, 2010). Similarly, gender-specific materials and policies like Girls' Day (www.girls-day.de), which has been held in Germany over twenty years, have also recently been suspected to foster gender stereotypes by reproducing them implicitly instead of reducing them (Wienkamp, 2018).

There is also a theoretical assumption pointing out the detrimental effects of stereotypes on performance. Stereotype threat theory (Steele & Aronson, 1995) argues that people suffer in their performance when negative stereotypes about their own group are salient. Steele and Aronson's original experiments showed that frequently observed SAT performance differences between African American and Caucasian American students decreased when demographics were not assessed until after the test. They suggested that reducing the relevance of ethnicity eliminates the salience of the negative stereotype about intellectual inferiority of African Americans, which otherwise decreased their performance. Based on this study, stereotype threat effects have been demonstrated in several domains and discriminated groups (for a review see Smith & Hung, 2008; Spencer et al., 2016). Besides ethnicity, gender has been the most frequently explored group category in stereotype threat research worldwide until today.

1.1 Effects of gender stereotypes on performance

Women and girls have been shown to be prone to stereotype threat in mathematical and spatial tasks (for a meta-analysis, Doyle & Voyer, 2016), natural sciences such as physics (Marchand & Taasoobshirazi, 2013) as well as information technology (Cooper, 2006). Taken together, stereotype threat appears to be a robust phenomenon with small effect sizes in girls $d=0.22$ (Flore & Wicherts, 2015) up to medium effect sizes in women $d=0.48$ (Walton & Spencer, 2009). Although mediating mechanisms of stereotype threat effects are still being discussed (Pennington et al., 2016), a working memory overload caused by a complex interaction of physiological, cognitive and affective processes (Schmader et al., 2008) is most commonly suggested and assured (Bedyńska et al., 2019).

While the first experiments regarding this phenomenon predominantly focused on adult women, usually students in laboratory contexts (for a meta-analysis see Nguyen & Ryan, 2008), stereotype threat was subsequently also shown in field research with children and adolescents (e.g. Hermann & Vollmeyer, 2017; Keller, 2007), even at primary school level (e.g. Hermann & Vollmeyer, 2016; Neuville & Croizet, 2007), down to the age of four (Shenouda & Danovitch, 2014). Furthermore, most studies on younger school girls used implicit methods to detect stereotype threat effects (for a review see Régner et al., 2014), proving gender stereotypes to become easily activated in learning environments. For example, it was shown

that simple situational cues like coloring stereotypical pictures (Neuville & Croizet, 2007), thinking about one's own gender (Ambady et al., 2001) or being confronted with the minority status of female mathematicians (Muzzatti & Agnoli, 2007) limited young girls' performance. Meanwhile, boys seemed to benefit from salient gender stereotypes in some stereotype threat experiments (Ambady et al., 2001; Neuburger et al., 2012) that highlighted supposed male superiority in mathematics and spatial abilities—a phenomenon already labeled stereotype lift (Walton & Cohen, 2003). Regarding this effect, meta-analysis revealed small ($d=0.24$, Walton & Cohen, 2003) but statistically significant performance increases under stereotype threat for indirectly upgraded persons. In contrast to negatively stereotyped girls, boys are supposed to be more self-confident in corresponding test situations due to downward comparison processes, facilitating their performance and potentially increasing the gender gap even more. Taken together, studies focusing straightly on stereotype lift in young boys seem to be rare, highlighting the importance of further research in this direction.

At least, studies with children have most often induced stereotype threat via implicit cues, for example by activating the gender category somehow, without stating anything explicitly about girls' or boys' performance. Whereas some studies instructed children to work with gender-specific task materials (coloring stereotypical pictures; Ambady et al., 2001; Hermann & Vollmeyer, 2016; Neuville & Croizet, 2007) others gave them questions about their gender (Ambady et al., 2001) or presented them a story about a stereotypically feminine girl (Tomasetto et al., 2011), before their performance was assessed. Furthermore, studies have simply varied the task description (reading task vs. a game) to manipulate a threat in boys, who are stereotyped to have lower reading abilities, what should be activated implicitly when their reading skills are at stake (Pansu et al., 2016). Others confronted primary school children predominantly male mathematicians (9 out of 10) to point on the female inferiority in the domain (Muzzatti & Agnoli, 2007).

1.2 Effects of gender stereotypes on motivation

Contrasting the negative effects of stereotypes on women's and girls' mathematical performance, their influence on underlying motivational processes seems to be much more complex. Whereas some authors suggest that motivation increases due to stereotype threat (Jamieson & Harkins, 2007), others suspect a motivational decrease (Shapiro & Williams, 2012; Thoman et al., 2013). On the one hand, there exists evidence that females under stereotype threat make a greater effort (Jamieson & Harkins, 2009; Seitchik & Harkins, 2015) as they are motivated to combat the negative expectations about their group and worried about mistakes (Brodish & Devine, 2009; Chalabaev et al., 2012; Smith, 2006). On the other hand, it was shown that negatively stereotyped females report lower self-efficacy (Cadaret et al., 2017; Deemer et al., 2014; Spencer et al., 1999), performance expectations (Cadinu et al., 2003; Smith, 2006), interest (Smith et al., 2007) and motivation to improve their skills (Fogliati & Bussey, 2013). Considering these multiple aspects, motivational

patterns under threat seem to be multifaceted and are still being discussed (for a review see Pennington et al., 2016).

In fact, some confusion about motivation under stereotype threat may be due to several methodological reasons. First, studies differed in their temporal perspective, as they focused immediate (Brodish & Devine, 2009; Cadinu et al., 2003; Chalabaev et al., 2012; Fogliati & Bussey, 2013; Jamieson & Harkins, 2007; Seitchik & Harkins, 2015; Smith et al., 2007) or long-time effects (Cardaret et al., 2017; Deemer et al., 2014; Thoman et al., 2013) of negative stereotyping. Second, some studies involved field research (Cardaret et al., 2017; Deemer et al., 2014), whereas others were run in the lab (Brodish & Devine, 2009; Chalabaev et al., 2012; Jamieson & Harkins, 2007; Seitchik & Harkins, 2015). Therefore, the salience of the negative stereotypes varied across studies, although it is questionable whether implicit threat cues (e.g. activating gender by assessing demographic data) and explicit stereotype threat activation (e.g. explicit statements about female mathematical inferiority) induce comparable reactions (e.g. Nguyen & Ryan, 2008). Finally, all of these studies on motivational change under stereotype threat explored adult women, making it important to find out how younger girls' motivation is shaped by gender stereotypes. As motivation under stereotype threat seems to be quite complex it might not be captured by a single motivational aspect. Therefore, we decided to run exploratory analyses based on the well-established self-determination theory (Deci & Ryan, 1985) which integrates several motivational aspects to explain the development of intrinsic motivation. Furthermore, intrinsic motivation was already defined as one central aspect of the stereotype threat process in the Motivational Experiences Model of Stereotype Threat (Thoman et al., 2013).

As stated by Deci and Ryan (1985, 1993), intrinsic motivation defined by the experience of interest and enjoyment is based on the fulfilment of three fundamental basic needs. First, people want to feel competent in their actions. Second, they want to have the feeling of acting autonomously instead of being under external pressure. Third, they want to feel connected with others and to belong in their social context. In a stereotype threat situation, all of these needs run the risk of being frustrated. Due to stereotype activation girls feel less competent in mathematics, they come under pressure to refute the negative expectations about their group and finally their belonging to their social context is called into question. Until now, stereotype threat research has predominantly focused on belonging uncertainty (Walton & Cohen, 2003), referred to the basic need of social belonging, which is disappointed when women enter STEM careers or contexts (for an overview see Thoman et al., 2013). However, most of these studies focused on adult women rather than younger girls. Immediate negative effects of stereotypes on the basic need of competence and autonomy, or rather perceived pressure, have not yet been tested explicitly in a stereotype threat situation specifically on younger girls. Although, Deci and Ryan (2000, 2002) postulated more proximal effects from these basic needs for the experience of situational intrinsic motivation, social belonging was seen as a more distal factor. In line with these assumptions in self-determination theory we wanted to explore how stereotypical mathematical tasks influence perceived interest and enjoyment as an immediate indicator for situational intrinsic motivation. Furthermore, situational effects on both basic needs will be tested by assessing children's perceived

competence and perceived pressure and tension while working with stereotypical tasks.

1.3 Research question and hypotheses

To sum up, in the present study we want to explore how gender-stereotypical tasks implicitly influence primary school children's performance and motivation in the classroom. For this purpose, we focus on potential negative effects of gender-stereotypical task materials in mathematics, originally intended to foster girls' motivation (e.g. where girls have to calculate ballet tickets). This is because large-scale assessment studies like PISA still show girls underperforming compared to boys in mathematics (OECD, 2016) although a lot of energy is invested in programs to foster girls' motivation in STEM. At the same time, detrimental effects of implicit activated stereotypes for girls' mathematical performance have been well documented (for a review see Régner et al., 2014) even though their effects on motivation are less well documented for young ages. Based on stereotype threat theory (Steele & Aronson, 1995) and in line with former stereotype threat studies with young children (Ambady et al., 2001; Hermann & Vollmeyer, 2016; Neuville & Croizet, 2007) the following hypotheses were investigated:

H1 Compared to gender-neutral mathematical tasks, we expect gender-stereotypical mathematical tasks to...

- (a) Decrease girls' performance (stereotype threat effect)
- (b) Increase boys' performance (stereotype lift effect).

Furthermore, we want to explore how gender-stereotypical tasks influence girls' and boys' motivational aspects, defined as interest and enjoyment, perceived competence and perceived pressure and tension based on self-determination theory (Deci & Ryan, 1985).

2 Method

2.1 Sample and design

Altogether $N=151$ primary school children (47.7% female; mean age: $M=9.81$, $SD=0.60$) participated in the study. They were attending fourth grade in three different schools in Germany. The collection of the data took place during a regular school session in a class context, with all children who were permitted to attend the test by their parents. The children were randomly assigned to one of two conditions as they either calculated gender-stereotypical or gender-neutral mathematical tasks. Hence, the study followed a 2 (sex: female vs. male) \times 2 (task: gender stereotypical vs. gender-neutral) design. Mathematical performance and motivational aspects were assessed as dependent variables.

Task 1: Pirates' prize

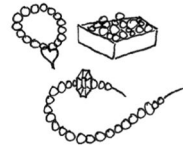
The pirates Pocke, Raffzahn and Knolle return from their raids. Pocke throws 5 pearl necklaces and 275 gold coins in the treasure chest, Raffzahn 3 sabers und 333 gold coins. Knolle could capture 4 golden rings and 286 gold coins.

How many gold coins have they captured together?

**Task 1: Crafting jewelry**

Katharina, Jaqueline and Maria want to make pearl jewelry together. Katharina brings 5 laces and 275 pearls, Jaqueline 3 bangles and 333 pearls with her. Maria brings some lace and 4 rings, 2 necklaces and 286 pearls with her.

How many pearls do they have together?

**Task 1: Carnival parade**

Hannah, Lukas and Kim were together on a carnival parade. Hannah caught 5 bags of popcorn and 275 hard candies, Lukas 3 lollipops and 333 hard candies. Kim could catch 4 bags of gummi bears and 286 hard candies.

How many sweets have they caught together?



Fig. 1 Implicit experimental manipulation. Note: Carnival celebrations include dressing up in costumes, dancing events and parades. Every town celebrating carnival boasts at least one parade with floats making fun of the themes of the day. Usually, sweets are thrown into the crowds lining the streets

2.2 Material and procedure

Based on the two gender-specific mathematics textbooks (Speicher, 2009a, b) we composed three different tests with mathematical word problems in which we manipulated gender stereotype salience implicitly. Whereas words, numbers and calculating operations were adjusted and held constant in all versions, design, pictures and the thematic embedment of the tasks was varied according to gender-stereotypical character (Themes of all tasks can be seen in Appendix A). For example, in the gender-stereotypical test version boys had to find out how many coins the pirates had captured, while girls calculated how many pearls were needed to make jewelry (see Fig. 1). Meanwhile, in the gender-neutral version students had to calculate the number of sweets collected at a carnival parade, which is traditionally a very popular event for children in Germany, like Halloween in the United States.

Overall, we selected six tasks which required different mathematical skills typically taught in the third and fourth grade in Germany (addition/subtraction up to 1000, multiplication tables, multiplication and division). Before testing the children, the tasks were checked by an expert teacher with respect to their difficulty. As the six tasks had several partial solutions every single result was rated 0 (incorrect) or 1 (correct), resulting in a score representing the percentage of correct outcomes. The reliability of the test was very good with Cronbach's $\alpha=0.85$. To complete the test children had 20 min. Childrens' motivational aspects were assessed afterwards with

Subscale	Sample item	Not at all	Totally	α
Interest/enjoyment	The tasks were exciting	○ ○ ○ ○ ○	○ ○ ○ ○ ○	.73
Perceived competence	I was good at these tasks	○ ○ ○ ○ ○	○ ○ ○ ○ ○	.80
Pressure/tension	I was often afraid to make mistakes	○ ○ ○ ○ ○	○ ○ ○ ○ ○	.62

Fig. 2 Sample items of the motivational subscales

Table 1 Observed and expected frequencies (in brackets) of the gender rated superior at mathematics, separated by experimental group

Group	Assigned gender			Total
	Boys	Girls	Both	
<i>Gender-stereotypical tasks</i>				
Girls	12 (18)	19 (12)	7 (8)	38
Boys	21 (18)	10 (12)	7 (8)	38
<i>Gender-neutral tasks</i>				
Girls	13 (15)	12 (10)	8 (7)	33
Boys	22 (17)	5 (12)	10 (8)	37

Data of $n=5$ children is missing resulting in a total $N=147$

a nine-item questionnaire, taken out of a German short version of the Intrinsic Motivation Inventory (Kurzskala Intrinsischer Motivation, KIM; Wilde et al., 2009) by Deci and Ryan (2003). According to self-determination theory (Deci & Ryan, 1985, 1993) the KIM contains four subscales to assess intrinsic motivation: interest and enjoyment, perceived competence, perceived choice as well as pressure and tension. Except for the perceived choice scale, which does not fit the test situation in our study, we used all subscales and adapted the item formulations as well as the design of the Likert scale to the younger age of our participants (see Fig. 2). Figure 2 also illustrates sample items for all subscales, each consisting of three items, showing acceptable to good reliability. All in all, the children had to answer nine items before giving their personal data (gender & age). To check our manipulation the children finally had to indicate whether they believed boys or girls to be better at mathematics. Here it was also allowed to mark both sexes if no gender difference was perceived. For a successful manipulation we expected boys to be seen as superior by all children who worked on stereotyped mathematical tasks, whereas no difference was expected in the control group.

3 Results

Before testing our hypotheses, we checked the success of our manipulation by conducting a Chi-Square test on the crosstab group x assigned competence. The results in Table 1 reveal significant group differences in the children's evaluation of girls'

Table 2 Intercorrelations of all dependent variables

	Mathematical performance	Interest and enjoyment	Perceived competence
Interest and enjoyment	0.05		
Perceived competence	0.37**	0.36**	
Pressure and tension	-0.30**	0.03	-0.40**

** Correlations significant at the 0.01 level (2-tailed)

Table 3 Means and standard deviations (in brackets) for all dependent variables

	Gender stereotypical tasks		Gender-neutral tasks	
	Girls ($n=39$)	Boys ($n=41$)	Girls ($n=33$)	Boys ($n=38$)
Mathematical performance	0.59 (0.26)	0.50 (0.22)	0.45 (0.20)	0.57 (0.24)
<i>Motivation</i>				
Interest and enjoyment	3.46 (0.98)	3.71 (0.96)	3.62 (0.83)	3.87 (0.80)
Perceived competence	3.50 (1.09)	3.67 (0.90)	3.63 (0.71)	3.99 (0.88)
Pressure and tension	2.96 (1.11)	2.58 (1.17)	2.95 (1.12)	2.54 (1.11)

and boys' mathematical abilities, $\chi^2(6, N=146)=13.58, p<0.05, \Phi=0.30$. However, contradicting our expectations, boys and girls assigned greater mathematical ability to their own gender after working on gender-stereotypical tasks. That means, the manipulation check turned out as expected only for boys.

As the different motivational aspects were not consistently correlated with each other and with performance, we ran separate ANOVAs with mathematical performance and motivational aspects, defined as interest and enjoyment, perceived competence as well as pressure and tension as dependent variables. Intercorrelations can be seen in Table 2.

The results regarding mathematical performance revealed a significant interaction between group and gender, $F(1,150)=7.32, p<0.01$. Surprisingly, the direction of this interaction was unexpected, as can be seen in Table 3: Whereas girls working on stereotypical mathematical tasks outperformed girls in the control group, boys' performance decreased while calculating the stereotypical tasks, compared to the boys in the control group. Nevertheless, focused contrast analysis showed that only the difference between the girls' test scores was significant, $t(147)=-2.40, p=0.02$. Therefore, Hypothesis 1a and 1b have to be rejected. Neither did girls experience a stereotype threat in their performance nor did boys a lift, due to the stereotypical character of the tasks. Instead, girls' performance increased significantly, whereas boys' performance did not drop.

Regarding the motivational aspects, we did not observe any main effects for the experimental group. In other words, children's estimated interest and enjoyment, perceived competence and perceived pressure and tension did not vary by the stereotypical design of the mathematical task. There was merely a significant main effect for gender as girls felt more pressure after the mathematical test in both groups,

$F(1,150)=4.56$, $p=0.03$. Girls were also less interested in the tasks and felt less competent while doing the test compared to the boys, independent of the experimental group. However, both main effects were slightly short of being significant and small in magnitude (interest and enjoyment, $F(1,150)=2.86$, $p=0.09$, $d=0.28$; perceived competence, $F(1,150)=3.19$, $p=0.08$, $d=0.29$).

4 Discussion

The present study explored the effect of gender-stereotypical tasks on children's performance and motivational aspects in mathematics in primary school. Based on the phenomenon of stereotype threat (Steele & Aronson, 1995) and stereotype lift (Walton & Cohen, 2003) we assumed gender-stereotypical tasks to result in decreased performance for girls (stereotype threat), while boys were supposed to excel in contrast to a control group working on gender-neutral mathematical tasks (stereotype lift). As effects of stereotypes on motivation have yielded mixed results in the past, various motivational aspects referring to self-determination theory (Deci & Ryan, 1985, 1993, 2002), i.e. interest and enjoyment, perceived competence as well as perceived pressure and tension were assessed for exploratory analysis. Contradicting our hypotheses, girls calculating stereotypical tasks outperformed girls in the control group, while boys' performance was slightly but not significantly lower than in the control group. Accordingly, the results contradicted our hypotheses, which therefore had to be rejected. Regarding motivation, only one significant gender difference appeared, as girls in both groups reported more pressure and tension while calculating the tasks, irrespective of the gender-stereotypical design. Theoretical and methodological explanations are discussed for girls' and boys' performance results separately before effects on motivation are addressed in a third section. Finally, a conclusion is derived, also containing implications for practice and further research.

4.1 Effects of stereotypical task performance

Results show that girls calculating gender-stereotypical tasks outperformed girls in the control group. This result contradicts our hypothesis and most stereotype threat studies in primary school children (for a review see Régner et al., 2014). However, studies have repeatedly failed to detect the detrimental effects (e.g. Flore et al., 2018), also in 10-year-old girls (Ambady et al., 2001) as in our study. Therefore, critical voices questioning the phenomenon of stereotype threat (Flore & Wicherts, 2015; Ganley et al., 2013; Stoet & Geary, 2012) have not become silent. Apart from that, these objections cannot explain the positive effects of stereotypical tasks on girls' mathematical performance we observed in our results. Therefore, three different explanations should be discussed below.

One reason for this effect could be that younger children temporarily see their own gender as superior, which is called in-group favoritism (Heyman & Legare, 2004), also fitting our manipulation check. In addition, female children working on stereotypical tasks more often assumed that girls were better at mathematics,

whereas choices in the control group were equally distributed for both genders. As stereotype threat predominantly operates on an unconscious level, it seems important to shed more light on studies exploring children's explicit and implicit maths stereotypes, where there have been mixed results in the past. Whereas some studies found traditional gender stereotypes favoring boys in mathematics on an explicit and implicit level in primary school girls (Cvencek et al., 2011), others found girls to see themselves as inferior only implicitly, while they explicitly assume they are excelling (Galdi et al., 2014; Passolunghi et al., 2014). Again, others found girls to see boys at an advantage in maths on an explicit level, whereas their implicit associations revealed that they saw themselves as superior (Steffens & Jelenec, 2011). However, last but not least results exist showing girls to feel superior implicitly and explicitly (Heyman & Legare, 2004; Nowicki & Lopata, 2017), contradicting traditional stereotypes altogether. Although explicit stereotype endorsement is not essential for stereotype threat to appear (Huguet & Régner, 2009; Spencer et al., 1999), the stereotype should at least be taken as valid (Jamieson & Harkins, 2010). In this regard, it was shown that girls indeed believed adult women to be inferior at mathematics, while they were not convinced that this stereotype held true for themselves (Martinot et al., 2012). Corresponding our results these authors also observed girls to perceive their mathematical performance as higher, when gender was made salient (Martinot & Désert, 2007). Therefore, we failed to detect stereotype threat effects on mathematical performance as girls of this age do not see themselves as negatively stereotyped either explicitly nor implicitly. Due to in-group favoritism the stereotyped tasks may heighten girls' self-assurance instead of harming their performance.

Another explanation could be that gender-stereotypical tasks facilitate girls' mathematical performance as they are less distracting, and instead more familiar. In line with this, studies exist on adults' reading performance and comprehension (Oakhill et al., 2005; Reynold et al., 2006), showing stereotyped content to be more easily processed, whereas counter-stereotypical content rather impeded performance due to higher cognitive load, harming working memory capacity. Correspondingly, it was recently shown in a primary school context that gender differences in mental rotation—stereotypically a boys' domain—disappeared when children had to work on gender-stereotyped objects (Ruthsatz et al., 2019). If children had to rotate a doll or a hair brush, girls performed as well as boys. Similar effects could also be detected when the traditional cube figures were changed into pellets (Ruthsatz et al., 2014). Again, the authors explained their results by girls' higher familiarity with handling the objects which could be more easily memorized by increasing the use of more holistic strategies (Ruthsatz et al., 2019), thus relieving working memory load. Paradoxically, limited working memory capacity has also been discussed as one central cognitive mediating mechanism in stereotype threat research (Schmader et al., 2008). However, which stereotypical content hampers or fosters working memory has not yet been systematically explored.

Therefore, it should at least also be considered that different gender-stereotypical cues could vary in their destructive potential. There are studies showing that particularly sexualized gender stereotypes have detrimental effects on girls' academic motivation (Brown, 2019), making it important to distinguish more between different categories of stereotype in future research. These sexualized gender stereotypes

teach girls to prioritize their physical attractiveness at the cost of other supposedly incompatible traits such as intelligence (Stone et al., 2015). Furthermore, it was also recently shown that the endorsement of these sexualized stereotypes was associated with lower academic outcomes even after controlling for general ability (Nelson & Brown, 2019). Similarly, primary school girls' mathematical performance decreased after they had been exposed to sexualized advertisements, compared to girls who saw non-sexualized materials (Pacilli et al., 2016). Therefore, it seems important to distinguish in future research between stereotyped content associated with sexualization, which potentially harms motivation and performance, and stereotypical cues which increase girls' familiarity and self-assurance when handling a task.

Regarding boys' mathematical performance in gender-stereotypical tasks, we did not observe a lift effect compared to the control group. Instead, boys in the control group tended to slightly outperform boys working on stereotypical tasks, although this trend did not reach significance. Contradicting stereotype lift theory (Walton & Cohen, 2003) and our hypothesis, this result is in line with other studies failing to detect stereotype lift effects in younger boys (e.g. Neuville & Croizet, 2007). Perhaps, another effect appeared for some male students, called choking under pressure (Baumeister, 1984)—a phenomenon which is defined as a performance drop people experience when they feel forced to fulfil extraordinarily high expectations of their group. According to this, boys have been more under pressure at school in general in the last decade (Kessels & Hannover, 2011). In this regard, it was also shown that boys experience stereotype threat due to their supposed academic shortfalls compared to girls (Hartley & Sutton, 2013). Boys even lost the stereotypically supposed advantages in their “favorite discipline” of mental rotation, when it was stated that girls achieved similar or even better results in the task (Neuburger et al., 2012) or stereotypically female objects had to be rotated (Ruthsatz et al., 2019). Indeed, boys showed faster rotation with stereotypically male objects (e.g. truck or gun), however, they also made more mistakes. Similar to girls, boys also suffer in their mathematical performance when exposed to gender stereotypes (Pacilli et al., 2016). However, irrespective of sexualization, most male gender stereotypes teach boys to be agentic, aggressive and dominant, all of which are characteristics associated with physical movement, maybe conflicting with boys' power of concentration in cognitive tasks. Therefore, future studies should explore in more detail how gender-stereotypical content and cues influence cognitive aspects, such as attention, distraction and concentration.

4.2 Effects of stereotypical tasks on motivational aspects

The results showed no significant differences in motivational aspects between the experimental conditions. Instead, the explored motivational aspects only varied by gender and not all differences reached significance. Whereas boys reported slightly but not significantly higher enjoyment and interest in the task as well as higher perceived competence, girls felt significantly more pressure and tension, independently of stereotypical task design. These results are in line with previous studies showing that girls experience more anxiety in mathematics (e.g. Erturan & Jansen, 2015)

while boys report more self-confidence, although girls receive comparable results (e.g. Hargreaves et al., 2008). However, similar to results showing intrinsic motivation failing to continuously predict performance throughout primary school (Garon-Carrier et al., 2016), the different aspects of intrinsic motivation explored in our study were also not consistently related to performance. That our stereotypical task design did not affect children's motivational aspects could at least also be due to a methodological reason. Bearing in mind the young age of our participants, it is also possible that the children thought that the motivational items referred to doing mathematics in general, instead of considering their stereotypical make up. Therefore, in a replication study—which is mandatory due to our hypothesis-contradicting results—it should be ensured that children keep the stereotypical task design in mind, while estimating motivational aspects.

4.3 Conclusion and implications

Taken together, our results neither revealed a destructive effect of stereotypical tasks on girls' mathematical performance, nor substantial advantages in performance for the boys. Contradicting our hypothesis, girls calculating stereotypical tasks outperformed girls in the control group, while boys' performance did not differ between both experimental conditions. Furthermore, for motivational aspects we merely found gender differences, confirming past results showing traditional gender disparities in mathematics, instead of effects of the stereotypical tasks. Regarding this, girls experienced significantly more pressure and tension while calculating, irrespective of the experimental condition. In line with the discussion about the nature of the stereotype threat, as a "cold" (cognitive) or "hot" (motivational) phenomenon (Schmader et al., 2008, p.348), performance-increasing effects of stereotypical tasks are more likely to stem from "cold" aspects in the shape of heightened familiarity-induced working memory relief than of advances in motivation. Otherwise, it is also possible that different stereotypical cues vary in their destructive potential, depending on whether they activate sexualized gender conceptions or just increase girls' familiarity and self-assurance in handling a task. Taking into account the numerous studies confirming implicit stereotype threat effects in the past (Régner et al., 2014), future studies should explore both, effects of stereotypes on children's cognitive processing and differences in their potential destructiveness. Therefore, maybe more qualitative research is needed, to enlighten children's perception of and associations with several stereotypical cues. In this regard, it would be also important to explore relevant moderating aspects, not yet considered in our study. Whether children's performance increases or decreases due to stereotypes could also depend on their parents' stereotypes (Tomasetto et al., 2011) or their implicit gender-math stereotypes, which have been shown to moderate stereotype threat and lift effects in female undergraduate students (Franceschini et al., 2014). However, studies focusing younger girls revealed explicit counter-stereotypical beliefs not to prevent stereotype threat (Huguet & Régner, 2009). Corresponding results would at least be important to develop safe interventions to combat stereotype threat, at which female role models have been shown to be helpful (for a review see Lawner et al., 2019)

although effect sizes vary substantially (for a meta-analysis see Liu et al., 2021). Thus, if “pink gives girls permission” to explore typical boy toys (Weisgram et al., 2014, p.401), a gender-stereotypical task design or role models may also function as an important gatekeeping step to foster girls’ mathematical performance and identification with STEM domains in the future.

Appendix A Task title, content and mathematical operation

Stereotypical tasks for the boys

- Pirates` Prize Addition of coins
- At the soccer stadium Subtraction of number of tickets for different matches
- A Magician Multiplication of lists of ingredients for a magic potion intended to make the magician stronger
- Trading cards (traditionally depicting soccer players in Germany) Multiplication and Division to determine the total number of cards or the number of packages respectively
- Badminton court Multiplication of the lengths of the sides of a badminton court to determine the area
- Spitting contest (cherry stones) Addition and division of distances

Stereotypical tasks for the girls

- Crafting jewelry Addition of pearls
- At the ballet Subtraction of numbers of tickets for different ballets (Swan Lake, Cinderella, Sleeping beauty, etc.)
- An ugly witch Multiplication of lists of ingredients for making a lemonade to become more attractive
- Naomi’s braids Multiplication and Division to determine the total number of strands or the number of braids respectively
- Street Art (Drawing a picture on the sidewalk) calculation of area using multiplication
- Crafting Garlands Calculating length using addition and division

Gender-neutral tasks

- Carnival parade Addition of sweets
- At the circus Subtraction of numbers of tickets
- A little vampire Multiplication of lists of ingredients to make a magic potion to allow for staying awake all day
- Lottery (a class won boxes with games) Multiplication and Division to determine the total number of games or the number of boxes respectively
- Chalk (Drawing a game on the sidewalk) calculation of area via multiplication

- The Move (Children got new wardrobes for their rooms) Calculating the wardrobes' widths using addition and division

Annotation: In the stereotypical tasks for boys and girls only characters of the corresponding gender appeared/were used. In contrast, the gender-neutral tasks, if groups of children were mentioned, they were always described as mixed gender groups.

Authors contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed and supervised by Johanna Maria Hermann and Regina Vollmeyer. The first draft of the manuscript was written by Johanna Maria Hermann and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. No funds, grants, or other support was received.

Data availability The data and materials can be obtained by emailing the first author.

Declarations

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

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the APA and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Consent to participate Written informed consent to participate in the study was obtained from one parent of each child.

Consent for publication Written informed consent for publication was included in the consent to participate.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

References

- Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science, 12*(5), 385–390. <https://doi.org/10.1111/1467-9280.00371>
- Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology, 46*(3), 610–620. <https://doi.org/10.1037/0022-3514.46.3.610>

- Bedyńska, S., Krejtz, I., & Sedek, G. (2019). Chronic stereotype threat and mathematical achievement in age cohorts of secondary school girls: Mediational role of working memory, and intellectual helplessness. *Social Psychology of Education, 22*(2), 321–335. <https://doi.org/10.1007/s11218-019-09478-6>
- Bigler, R. S., & Signorella, M. L. (2011). Single-sex education: New perspectives and evidence on a continuing controversy. *Sex Roles, 65*(9–10), 659–669. <https://doi.org/10.1007/s11199-011-0046-x>
- Brodish, A. B., & Devine, P. G. (2009). The role of performance–avoidance goals and worry in mediating the relationship between stereotype threat and performance. *Journal of Experimental Social Psychology, 45*(1), 180–185. <https://doi.org/10.1016/j.jesp.2008.08.005>
- Brown, C. S. (2019). Sexualized gender stereotypes predict girls' academic self-efficacy and motivation across middle school. *International Journal of Behavioral Development, 43*(6), 523–529. <https://doi.org/10.1177/0165025419862361>
- Budde, J., Kansteiner, K., & Bossen, A. (2016). *Zwischen Differenz und Differenzierung: Erziehungswissenschaftliche Forschung zu Mono- und Koedukation [Between difference and differentiation: Educational research about single- and coeducation]*. Springer.
- Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology, 33*(2), 267–285. <https://doi.org/10.1002/ejsp.145>
- Cadaret, M. C., Hartung, P. J., Subich, L. M., & Weigold, I. K. (2017). Stereotype threat as a barrier to women entering engineering careers. *Journal of Vocational Behavior, 99*, 40–51. <https://doi.org/10.1016/j.jvb.2016.12.002>
- Chalabaev, A., Major, B., Sarrazin, P., & Cury, F. (2012). When avoiding failure improves performance: Stereotype threat and the impact of performance goals. *Motivation and Emotion, 36*(2), 130–142. <https://doi.org/10.1007/s11031-011-9241-x>
- Cooper, J. (2006). The digital divide: The special case of gender. *Journal of Computer Assisted Learning, 22*, 320–334. <https://doi.org/10.1111/j.1365-2729.2006.00185.x>
- Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Development, 82*(3), 766–779. <https://doi.org/10.1111/j.1467-8624.2010.01529.x>
- Datnow, A., Hubbard, L., & Woody, E. (2001). Is single gender schooling viable in the Public Sector? Lessons from California's Pilot Program. Final Report. Toronto, Canada: Ontario Institute for Studies in Education, University of Toronto. Retrieved from <http://files.eric.ed.gov/fulltext/ED471051.pdf>
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum.
- Deci, E. L., & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik [Self-determination theory of motivation and its significance for pedagogy]. *Zeitschrift Für Pädagogik, 39*, 223–238.
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*, 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- Deci, E. L., & Ryan, R. M. (Eds.). (2002). *Handbook of self-determination research*. The University of Rochester Press.
- Deci, E. L., & Ryan, R. M. (2003). *Intrinsic Motivation Inventory*. Retrieved from <http://www.selfdeterminationtheory.org/intrinsic-motivation-inventory/>
- Deemer, E. D., Thoman, D. B., Chase, J. P., & Smith, J. L. (2014). Feeling the threat: Stereotype threat as a contextual barrier to women's science career choice intentions. *Journal of Career Development, 41*(2), 141–158. <https://doi.org/10.1177/0894845313483003>
- Doyle, R. A., & Voyer, D. (2016). Stereotype manipulation effects on math and spatial test performance: A meta-analysis. *Learning and Individual Differences, 47*, 103–116. <https://doi.org/10.1016/j.lindif.2015.12.018>
- Erturan, S., & Jansen, B. (2015). An investigation of boys' and girls' emotional experience of math, their math performance, and the relation between these variables. *European Journal of Psychology of Education, 30*(4), 421–435. <https://doi.org/10.1007/s10212-015-0248-7>
- Fabes, R. A., Pahlke, E., Martin, C. L., & Hanish, L. D. (2013). Gender-segregated schooling and gender stereotyping. *Educational Studies, 39*(3), 315–319. <https://doi.org/10.1080/03055698.2012.760442>

- Fabes, R. A., Martin, C. L., Hanish, L. D., Galligan, K., & Pahlke, E. (2015). Gender-segregated schooling: A problem disguised as a solution. *Educational Policy*, 29(3), 431–447. <https://doi.org/10.1177/0895904813492382>
- Faulstich-Wieland, H. (2011). Koedukation – Monoedukation [Regular education – single-sex education]. In H. Faulstich-Wieland, (Eds.): *Enzyklopädie Erziehungswissenschaft Online. Fachgebiet: Geschlechterforschung [Encyclopedia educational science online. Field: Gender studies]*. Juventa.
- Flore, P. C., Mulder, J., & Wicherts, J. M. (2018). The influence of gender stereotype threat on mathematics test scores of Dutch high school students: A registered report. *Comprehensive Results in Social Psychology*, 3(2), 140–174. <https://doi.org/10.1080/23743603.2018.1559647>
- Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A Meta-Analysis. *Journal of School Psychology*, 53(1), 25–44. <https://doi.org/10.1016/j.jsp.2014.10.002>
- Fogliati, V. J., & Bussey, K. (2013). Stereotype threat reduces motivation to improve: Effects of stereotype threat and feedback on women's intentions to improve mathematical ability. *Psychology of Women Quarterly*, 37(3), 310–324. <https://doi.org/10.1177/0361684313480045>
- Franceschini, G., Galli, S., Chiesi, F., & Primi, C. (2014). Implicit gender–math stereotype and women's susceptibility to stereotype threat and stereotype lift. *Learning and Individual Differences*, 32, 273–277. <https://doi.org/10.1016/j.lindif.2014.03.020>
- Galdi, S., Cadinu, M., & Tomasetto, C. (2014). The roots of stereotype threat: When automatic associations disrupt girls' math performance. *Child Development*, 85(1), 250–263. <https://doi.org/10.1111/cdev.12128>
- Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An examination of stereotype threat effects on girls' mathematics performance. *Developmental Psychology*, 49(10), 1886–1897. <https://doi.org/10.1037/a0031412>
- Garon-Carrier, G., Boivin, M., Guay, F., Kovas, Y., Dionne, G., Lemelin, J. P., Séguan, J. R., Vitaro, F., & Tremblay, R. E. (2016). Intrinsic motivation and achievement in mathematics in elementary school: A longitudinal investigation of their association. *Child Development*, 87(1), 165–175. <https://doi.org/10.1111/cdev.12458>
- Guggenbühl, A. (2008): Die Schule – ein weibliches Biotop [School – a female habitat] ? In: M. Matzner & W. Tischner (Eds.): *Handbuch Jungen-Pädagogik [Handbook boys' education]* (pp. 150–169). Beltz,
- Gurian, M., Henley, P., & Trueman, T. (2001). *Boys and girls learn differently!* Jossey-Bass.
- Hannover, B., & Kessels, U. (2002). Monoedukativer Anfangsunterricht in Physik in der Gesamtschule: Auswirkungen auf Motivation, Selbstkonzept und Einteilung in Grund- oder Fortgeschrittenenkurse [Physics in single-sex or mixed teaching groups. Impact on motivation, self-concept of abilities, and course enrollment of 8th grade comprehensive school students]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 34(4): 201–215. doi: <https://doi.org/10.1026/0049-8637.34.4.201>
- Hannover, B., & Kessels, U. (2011). Sind Jungen die neuen Bildungsverlierer? Empirische Evidenz für Geschlechterdisparitäten zuungunsten von Jungen und Erklärungsansätze [Are boys left behind at school? Reviewing and explaining education-related gender disparities]. *Zeitschrift Für Pädagogische Psychologie*, 25, 89–103. <https://doi.org/10.1024/1010-0652/a000039>
- Häussler, P., & Hoffmann, L. (2002). An intervention study to enhance girls' interest, self-concept, and achievement in physics classes. *Journal of Research in Science Teaching*, 39(9), 870–888. <https://doi.org/10.1002/tea.10048>
- Hargreaves, M., Homer, M., & Swinnerton, B. (2008). A comparison of performance and attitudes in mathematics amongst the 'gifted'. Are boys better at mathematics or do they just think they are? *Assessment in Education: Principles, Policy & Practice*, 15(1), 19–38. <https://doi.org/10.1080/09695940701876037>
- Hartley, B. L., & Sutton, R. M. (2013). A stereotype threat account of boys' academic underachievement. *Child Development*, 84(5), 1716–1733. <https://doi.org/10.1111/cdev.12079>
- Hermann, J. M., & Vollmeyer, R. (2016). Stereotype Threat in der Grundschule [Stereotype threat in primary school]. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, 48, 42–49. <https://doi.org/10.1026/0049-8637/a000143>
- Hermann, J. M., & Vollmeyer, R. (2017). Das mathematische Selbstkonzept als Moderator des Stereotype-Threat-und Stereotype-Lift-Effekts [The mathematical self-concept as moderator of the stereotype threat and stereotype lift effects]. *Zeitschrift Für Pädagogische Psychologie*, 31, 211–234. <https://doi.org/10.1024/1010-0652/a000209>

- Heyder, A., & Kessels, U. (2013). Is school feminine? Implicit gender stereotyping of school as a predictor of academic achievement. *Sex Roles, 69*(11–12), 605–617. <https://doi.org/10.1007/s11199-013-0309-9>
- Heyman, G. D., & Legare, C. H. (2004). Children's beliefs about gender differences in the academic and social domains. *Sex Roles, 50*(3–4), 227–239. <https://doi.org/10.1023/B:SERS.0000015554.12336.30>
- Hilliard, L. J., & Liben, L. S. (2010). Differing levels of gender salience in preschool classrooms: Effects on children's gender attitudes and intergroup bias. *Child Development, 81*(6), 1787–1798. <https://doi.org/10.1111/j.1467-8624.2010.01510.x>
- Huguet, P., & Régner, I. (2009). Counter-stereotypic beliefs in math do not protect school girls from stereotype threat. *Journal of Experimental Social Psychology, 45*(4), 1024–1027. <https://doi.org/10.1016/j.jesp.2009.04.029>
- Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology, 93*(4), 544–564. <https://doi.org/10.1037/0022-3514.93.4.544>
- Jamieson, J. P., & Harkins, S. G. (2009). The effect of stereotype threat on the solving of quantitative GRE problems: A mere effort interpretation. *Personality and Social Psychology Bulletin, 35*(10), 1301–1314. <https://doi.org/10.1177/0146167209335165>
- Jamieson, J. P., & Harkins, S. G. (2010). Evaluation is necessary to produce stereotype threat performance effects. *Social Influence, 5*(2), 75–86. <https://doi.org/10.1080/15534510903512409>
- Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *British Journal of Educational Psychology, 77*(2), 323–338. <https://doi.org/10.1348/000709906X113662>
- Kleinfeld, J. (2009). The state of American boyhood. *Gender Issues, 26*(2), 113–129. <https://doi.org/10.1007/s12147-009-9074-z>
- Lawner, E. K., Quinn, D. M., Camacho, G., Johnson, B. T., & Pan-Weisz, B. (2019). Ingroup role models and underrepresented students' performance and interest in STEM: A meta-analysis of lab and field studies. *Social Psychology of Education, 22*(5), 1169–1195. <https://doi.org/10.1007/s11218-019-09518-1>
- Liu, S., Liu, P., Wang, M., & Zhang, B. (2021). Effectiveness of stereotype threat interventions: A meta-analytic review. *Journal of Applied Psychology, 106*(6), 921–949. <https://doi.org/10.1037/ap10000770>
- MacPhee, D., & Prendergast, S. (2019). Room for improvement: Girls' and boys' home environments are still gendered. *Sex Roles, 80*(5–6), 332–346. <https://doi.org/10.1007/s11199-018-0936-2>
- Marchand, G. C., & Taasobshirazi, G. (2013). Stereotype threat and women's performance in physics. *International Journal of Science Education, 35*, 3050–3061. <https://doi.org/10.1080/09500693.2012.683461>
- Martinot, D., Bagès, C., & Désert, M. (2012). French children's awareness of gender stereotypes about mathematics and reading: When girls improve their reputation in math. *Sex Roles, 66*(3–4), 210–219. <https://doi.org/10.1007/s11199-011-0032-3>
- Martinot, D., & Désert, M. (2007). Awareness of a gender stereotype, personal beliefs and self-perceptions regarding math ability: When boys do not surpass girls. *Social Psychology of Education, 10*(4), 455–471. <https://doi.org/10.1007/s11218-007-9028-9>
- Murnen, S. K., Greenfield, C., Younger, A., & Boyd, H. (2016). Boys act and girls appear: A content analysis of gender stereotypes associated with characters in children's popular culture. *Sex Roles, 74*(1–2), 78–91. <https://doi.org/10.1007/s11199-015-0558-x>
- Muzzatti, B., & Agnoli, F. (2007). Gender and mathematics: Attitudes and stereotype threat susceptibility in Italian children. *Developmental Psychology, 43*(3), 747–759. <https://doi.org/10.1037/0012-1649.43.3.747>
- Nelson, A. A., & Brown, C. S. (2019). Too pretty for homework: Sexualized gender stereotypes predict academic attitudes for gender-typical early adolescent girls. *The Journal of Early Adolescence, 39*(4), 603–617. <https://doi.org/10.1177/0272431618776132>
- Neuburger, S., Jansen, P., Heil, M., & Quaiser-Pohl, C. (2012). A threat in the classroom: Gender stereotype activation and mental-rotation performance in elementary-school children. *Zeitschrift Für Psychologie, 220*(2), 61–69. <https://doi.org/10.1027/2151-2604/a000097>
- Neuville, E., & Croizet, J. C. (2007). Can salience of gender identity impair math performance among 7–8 years old girls? The moderating role of task difficulty. *European Journal of Psychology of Education, 22*(3), 307–316. <https://doi.org/10.1007/BF03173428>

- Nguyen, H. H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology, 93*, 1314–1334. <https://doi.org/10.1037/a0012702>
- Nowicki, E. A., & Lopata, J. (2017). Children's implicit and explicit gender stereotypes about mathematics and reading ability. *Social Psychology of Education, 20*(2), 329–345. <https://doi.org/10.1007/s11218-015-9313-y>
- Oakhill, J., Garnham, A., & Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Memory & Cognition, 33*(6), 972–983. <https://doi.org/10.3758/BF03193206>
- OECD (2016). PISA 2015 Results (Volume I): *Excellence and Equity in Education*, PISA, OECD Publishing. <http://dx.doi.org/https://doi.org/10.1787/9789264266490-en>
- Pacilli, M. G., Tomasetto, C., & Cadinu, M. (2016). Exposure to sexualized advertisements disrupts children's math performance by reducing working memory. *Sex Roles, 74*(9–10), 389–398. <https://doi.org/10.1007/s11199-016-0581-6>
- Pahlke, E., Hyde, J. S., & Allison, C. M. (2014). The effects of single-sex compared with coeducational schooling on students' performance and attitudes: A meta-analysis. *Psychological Bulletin, 140*(4), 1042–1072. <https://doi.org/10.1037/a0035740>
- Pahlke, E., & Hyde, J. S. (2016). The debate over single-sex schooling. *Child Development Perspectives, 10*(2), 81–86. <https://doi.org/10.1111/cdep.12167>
- Pansu, P., Régner, I., Max, S., Colé, P., Nezelek, J. B., & Huguet, P. (2016). A burden for the boys: Evidence of stereotype threat in boys' reading performance. *Journal of Experimental Social Psychology, 65*, 26–30. <https://doi.org/10.1016/j.jesp.2016.02.008>
- Passolunghi, M. C., Ferreira, T. I. R., & Tomasetto, C. (2014). Math-gender stereotypes and math-related beliefs in childhood and early adolescence. *Learning and Individual Differences, 34*, 70–76. <https://doi.org/10.1016/j.lindif.2014.05.005>
- Pennington, C. R., Heim, D., Levy, A. R., & Larkin, D. T. (2016). Twenty years of stereotype threat research: A review of psychological mediators. *PloS One, 11*(1), e0146487. <https://doi.org/10.1371/journal.pone.0146487>
- Régner, I., Steele, J. R., Ambady, N., Thinus-Blanc, C., & Huguet, P. (2014). Our future scientists: A review of stereotype threat in girls from early elementary school to middle school. *Revue Internationale De Psychologie Sociale, 27*(3), 13–51.
- Reynolds, D., Garnham, A., & Oakhill, J. (2006). Evidence of immediate activation of gender information from a social role name. *Quarterly Journal of Experimental Psychology, 59*(5), 886–903. <https://doi.org/10.1080/02724980543000088>
- Rudolph-Albert, F., & Keller, J. (2007). Mathematikunterricht in monoedukativen und koedukativen Schulen – Lernen mit heuristischen Lösungsbeispielen [mathematic education in single-sex and regular schools – Learning with heuristic solution examples]. In: L. Herwartz-Emden (Eds.). *Neues aus alten Schulen. Empirische Studien in Mädchenschulen [News from old schools. Empirical studies in girl schools]*. pp. 181–203. Budrich,
- Ruthsatz, V., Neuburger, S., Jansen, P., & Quaiser-Pohl, C. (2014, September). Pellet figures, the feminine answer to cube figures? Influence of stimulus features and rotational axis on the mental-rotation performance of fourth-grade boys and girls. In *International Conference on Spatial Cognition* (pp. 370–382). Springer, Cham. doi: https://doi.org/10.1007/978-3-319-11215-2_26
- Ruthsatz, V., Rahe, M., Schürmann, L., & Quaiser-Pohl, C. (2019). Girls' stuff, boys' stuff and mental rotation: Fourth graders rotate faster with gender-congruent stimuli. *Journal of Cognitive Psychology, 31*(2), 225–239. <https://doi.org/10.1080/20445911.2019.1567518>
- Salomone, R. C. (2004). Feminist voices in the debate over single-sex schooling: Finding common ground. *Michigan Journal of Gender and Law, 11*, 63–95.
- Sax, L. (2006). Six degrees of separation: What teachers need to know about the emerging science of sex differences. *Educational Horizons, 84*(3), 190–200. <https://www.jstor.org/stable/42926590>
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*(2), 336–356. <https://doi.org/10.1037/0033-295X.115.2.336>
- Seitchik, A. E., & Harkins, S. G. (2015). Stereotype threat, mental arithmetic, and the mere effort account. *Journal of Experimental Social Psychology, 61*, 19–30. <https://doi.org/10.1016/j.jesp.2015.06.006>
- 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>

- Shenouda, C. K., & Danovitch, J. H. (2014). Effects of gender stereotypes and stereotype threat on children's performance on a spatial task. *Revue Internationale De Psychologie Sociale*, 27(3), 53–77.
- Smith, J. L. (2006). The interplay among stereotypes, performance-avoidance goals, and women's math performance expectations. *Sex Roles*, 54(3–4), 287–296. <https://doi.org/10.1007/s11199-006-9345-z>
- Smith, C. S., & Hung, L. C. (2008). Stereotype threat: Effects on education. *Social Psychology of Education*, 11(3), 243–257. <https://doi.org/10.1007/s11218-008-9053-3>
- Smith, J. L., Sansone, C., & White, P. H. (2007). The stereotyped task engagement process: The role of interest and achievement motivation. *Journal of Educational Psychology*, 99(1), 99. <https://doi.org/10.1037/0022-0663.99.1.99>
- Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology*, 67, 415–437. <https://doi.org/10.1146/annurev-psych-073115-103235>
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4–28. <https://doi.org/10.1006/jesp.1998.1373>
- Speicher, K. (2009a). *Textaufgaben für Mädchen: 100 Aufgaben, die Mädchen wirklich begeistern. 2. bis 4. Klasse [Mathematical word problems for girls. 100 tasks that really engage girls' interest]*. PONS.
- Speicher, K. (2009b). *Textaufgaben für Jungs: 100 Aufgaben, die Jungs wirklich begeistern. 2. bis 4. Klasse [Mathematical word problems for boys. 100 tasks that really engage boys' interest]*. PONS.
- Spinner, L., Cameron, L., & Calogero, R. (2018). Peer toy play as a gateway to children's gender flexibility: The effect of (counter) stereotypic portrayals of peers in children's magazines. *Sex Roles*, 79(5–6), 314–328. <https://doi.org/10.1007/s11199-017-0883-3>
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811. <https://doi.org/10.1037/0022-3514.69.5.797>
- Steffens, M. C., & Jelenec, P. (2011). Separating implicit gender stereotypes regarding math and language: Implicit ability stereotypes are self-serving for boys and men, but not for girls and women. *Sex Roles*, 64(5–6), 324–335. <https://doi.org/10.1007/s11199-010-9924-x>
- Stoet, G., & Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement? *Review of General Psychology*, 16(1), 93–102. <https://doi.org/10.1037/a0026617>
- Stone, E. A., Brown, C. S., & Jewell, J. A. (2015). The sexualized girl: A within-gender stereotype among elementary school children. *Child Development*, 86(5), 1604–1622. <https://doi.org/10.1111/cdev.12405>
- Thoman, D. B., Smith, J. L., Brown, E. R., Chase, J., & Lee, J. Y. K. (2013). Beyond performance: A motivational experiences model of stereotype threat. *Educational Psychology Review*, 25(2), 211–243. <https://doi.org/10.1007/s10648-013-9219-1>
- Tomasetto, C., Alparone, F. R., & Cadinu, M. (2011). Girls' math performance under stereotype threat: The moderating role of mothers' gender stereotypes. *Developmental Psychology*, 47(4), 943–949. <https://doi.org/10.1037/a0024047>
- Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, 39(5), 456–467. [https://doi.org/10.1016/S0022-1031\(03\)00019-2](https://doi.org/10.1016/S0022-1031(03)00019-2)
- Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science*, 20(9), 1132–1139. <https://doi.org/10.1111/j.1467-9280.2009.02417.x>
- Weisgram, E. S., Fulcher, M., & Dinella, L. M. (2014). Pink gives girls permission: Exploring the roles of explicit gender labels and gender-typed colors on preschool children's toy preferences. *Journal of Applied Developmental Psychology*, 35(5), 401–440. <https://doi.org/10.1016/j.appdev.2014.06.004>
- Wienkamp, G. (2018). Der Girls' Day aus geschlechtersoziologischer Perspektive. Wie das monoedukative Förderprogramm Geschlechterdifferenzen und-stereotype reproduzieren und neutralisieren kann [The Girls' Day under gender sociological perspective. How single sex educational promotion programs reproduce and neutralizes gender differences and stereotypes]. *IZGOnZeit. Onlinezeitschrift des Interdisziplinären Zentrums für Geschlechterforschung (IZG)*, 30–43. doi: <https://doi.org/10.4119/UNIBI/izgonzeit-109>

Wilde, M., Bätz, K., Kovaleva, A., & Urhahne, D. (2009). Überprüfung einer Kurzsкала intrinsischer Motivation (KIM) [Testing a short scale of intrinsic motivation]. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 31–45. <https://pub.uni-bielefeld.de/record/2404161>

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

Johanna Maria Hermann is a postdoctoral research fellow at the Department of Educational Psychology at the Goethe University Frankfurt, Germany. Her research interests include effects of stereotypes on educational and sports performance as well as motivation in performance and learning processes.

Regina Vollmeyer is a full professor at the Department of Educational Psychology, Goethe University Frankfurt, Germany. Her research interests focus on experimental research on motivation in performance and learning processes as well as problem solving.