

**Solving the Gender Equation?  
Measuring and Refuting Teachers'  
Math-Gender Misconceptions**

Anna-Sophia Dersch

April 2025

Inaugural Dissertation  
for the Attainment of a Doctoral Degree in the Natural Sciences

Submitted to the Faculty of Psychology and Sports Science  
Justus-Liebig-University Giessen

First Thesis Reviewer: Prof. Dr. Alexander Eitel

Second Thesis Reviewer: Prof. Dr. Lisa Bardach

Dean: Prof. Dr. Christiane Hermann

## Acknowledgments

I want to thank everyone who has contributed to my research in this dissertation and beyond. Thank you to my students for their interest and dedication. Teaching you and supervising your theses has advanced my insights about conceptual change, mathematics and gender, and psychological misconceptions. Thank you to the reliable research assistants and interns. Thank you to the pre-service and in-service teachers who participated in my studies.

I further want to thank my co-authors, Dr. Ian Thacker and Prof. Dr. Anke Heyder, who have worked with me on the manuscripts included in this dissertation. A special thank you goes to Anke, who has accompanied me with helpful advice almost from the beginning of my PhD.

Furthermore, I would like to thank the researchers whose labs I had the honor of visiting during my PhD. Prof. Dr. Ladislao Salmeron, thank you for welcoming me to your lab in Valencia and promoting my collaboration with your brilliant team. Prof. Dr. Helen Watt, thank you for introducing me to many potential collaborators in Sydney and beyond; with your help, I established many valuable connections in gender and STEM research. Prof. Dr. Stella Vosniadou and Prof. Dr. Mike Lawson, thank you for warmly welcoming me to Adelaide and challenging my ideas in a way that has helped me progress as a researcher.

I would also like to thank the members of my defense committee. Thank you to Prof. Dr. Jan Häusser for co-supervising my dissertation from a social psychology perspective. Thank you to Prof. Dr. Lisa Bardach, who has agreed to dedicate valuable time to reading and evaluating this dissertation. Thank you to Prof. Dr. Claudia von Aufschnaiter for being part of my dissertation defense committee and providing her expertise on conceptual change, being a successful woman in STEM herself.

Most importantly, I want to thank my supervisor, Prof. Dr. Alexander Eitel, with whom I have enjoyed working for many years, first as a research assistant in Freiburg and then as a doctoral student in Gießen. Thank you for taking my ideas and perspectives seriously. I have always felt very appreciated working with you. Thank you for providing me with great opportunities. Thank you for letting me explore my own ways while being supportive when needed. Thank you for being a role model on how to be a great researcher—even if I still have a way to go.

I want to thank everyone who has accompanied me during my PhD. Thank you to my colleagues in Giessen for the fruitful discussions, constructive support in research questions, and the lunch and coffee breaks in a pleasant environment, making long office hours bearable. I also want to thank my collaborators for warmly welcoming me to São Tomé, Adelaide, Sydney, and Valencia.

A fond thank you goes to all the inspiring young researchers I have met and become friends with along the way. It has been encouraging and comforting to share some of the unique privileges and struggles of being a young researcher with you. Thank you for making the conferences we attended memorable and for providing a supportive network.

Last but most definitely not least, I would like to thank my wonderful friends and family, who provided emotional support whenever I needed it, took my mind off things at times, showed interest in my research, and always believed in me. A special thank you goes to my partner, Yves, for always being there and experiencing all the highs and lows with me.

## Contents of this Dissertation

### Doctoral Thesis

Introduction .....	1
The Math-Gender Stereotype .....	1
Teachers' Math-Gender Stereotype-Reinforcing Behaviors .....	2
Math-Gender Misconceptions .....	4
Conceptual Change Using Refutation Texts .....	4
Challenges for Teachers' Conceptual Change .....	6
Pitfalls and Potentials of Targeting Math-Gender Misconceptions .....	7
Goals of the Present Dissertation.....	9
Summaries of Studies Included in this Dissertation .....	10
Study 1 .....	10
Study 2 .....	11
Study 3 .....	13
General Discussion .....	15
Pattern of Results of this Dissertation .....	16
Addressing Gaps from Previous Research .....	16
Limitations and Further Research .....	24
Conclusion and Implications for Practice .....	27
References .....	28
<i>Study 1. Exploring the Nature of Teachers' Math-Gender Stereotypes:</i> The Math-Gender Misconception Questionnaire.....	40
<i>Study 2. Teachers' Awareness of Math-gender Stereotype-reinforcing Behaviors in the Classroom:</i> A Video Experiment .....	54
<i>Study 3. Should We Tell Teachers That They Are (In-)Correct?</i> How Personalized Refutation Text Promotes or Hampers Conceptual Change .....	99
List of Publications .....	143
Declaration of Originality .....	144

## Introduction

### The Math-Gender Stereotype

Women are currently underrepresented in *science, technology, engineering, and mathematical* (STEM) fields (Avolio et al., 2020; Wang & Degol, 2017), especially in math-intensive STEM fields and mathematics (Kahn & Ginther, 2017). This female underrepresentation in mathematics and math-intensive STEM fields manifests in the *math-gender gap*. The reasons for the math-gender gap are multifactorial and complex (e.g., Avolio et al., 2020). These multiple factors can hardly be researched separately from one another since they interact from birth (Halpern et al., 2007; Wang & Degol, 2017). Nevertheless, empirical evidence for social influences on the underrepresentation of women in math-intensive STEM fields is abundant (e.g., Avolio et al., 2020; Wang & Degol, 2017). One such social factor influencing this underrepresentation is the *math-gender stereotype* (e.g., Cvencek et al., 2012; Eccles, 2011; Girelli, 2023; Nosek & Smyth, 2011; Steffens & Jelenec, 2011).

There are different definitions of stereotypes based on different theoretical assumptions about how our knowledge and beliefs are stored and encoded. Many definitions agree that stereotypes are generalized, socially shared sets of knowledge and beliefs about members of a social group and the social group's (assumed) characteristics (e.g., Hilton & von Hippel, 1996; Klineberg, 1951; Vinacke, 1957). Based on this generalized knowledge and beliefs, stereotypes, as the cognitive component of an attitude, allow for the evaluation of social groups (Fiske, 1982). The goal of stereotypes is to reduce the complexity of one's social environment, making it appear more predictable and controllable (e.g., Bodenhausen, 1990; Fiske, 1982; Hilton & von Hippel, 1996). According to schema theories, stereotypes are stored in long-term memory, clustered with all the knowledge about the social group they refer to (e.g., Martin & Halverson, 1981). Stereotypes are often learned and consolidated from early childhood (Hamilton & Sherman, 2014; Hinton, 2017). The same applies to gender stereotypes about (assumed) differences between gender groups (Ellemers, 2018). Most Western societies' salient binary categorization into female and male contributes to developing such gender stereotypes (Ellemers, 2018; Martin & Halverson, 1981). The math-gender stereotype is a set of beliefs connoting mathematics as male or the mere association of the categories *mathematics* and *male* (Ashmore & DelBoca, 1979; Greenwald et al., 2002).

On the one hand, stereotypes are shaped by social reality (Jussim et al., 2005). The lower representation of women in mathematics and math-intensive STEM fields (Kahn & Ginther, 2017; Miller et al., 2015) is reflected in the math-gender stereotype (Miller et al., 2015). On the other hand, stereotypes shape social reality (Eagly & Koenig, 2021). The math-gender stereotype, therefore, predicts lower female representation in mathematics and math-intensive STEM fields (e.g., Hand et al., 2017; Halpern et al., 2007; Nosek & Symth, 2011). This influence can occur via different mechanisms.

*Balanced Identity Theory* (Cvencek et al., 2012; Greenwald et al., 2002) describes how implicit learning perpetuates the math-gender stereotype. People socialized as female identify with the category *female* and, due to common binary gender categorizations, disidentify with the category *male*. Thus, these girls and women also disidentify with mathematics because of the association of *mathematics* with *male*, hindering the association of the own female self with mathematics (Cvencek et al., 2012; Greenwald et al., 2002).

The *Situated Model of Achievement Related Choices* (Eccles & Wigfield, 2024) focuses on the motivational influences of the math-gender stereotype. It proposes that the math-gender stereotype contributes to lowering female learners' expectation of success and female learners' subjective task value in mathematics (Eccles, 1994; 2011; Song et al., 2017). The reduction of subjective task value is most prominent for intrinsic value, meaning the subjective enjoyment of mathematics (Watt, 2004; Watt et al., 2012). Both an overall lowered subjective task value and a

lowered expectation of success influence female learners' achievement-related choices (Eccles, 2011; Eccles & Wigfield, 2024). Female learners' lowered subjective task value in mathematics, for instance, has been found to partially explain gender differences in STEM career attainment (Wang et al., 2015). Additionally, subjective task value in female learners mediates the relationship between the math-gender stereotype and female learners' educational- and career choices against mathematical fields (Chow & Salmela-Aro, 2011; Eccles et al., 1999; Nagy et al., 2008; Song et al., 2017). In sum, the math-gender stereotype lowers female representation in mathematics (e.g., Eccles, 2011; Greenwald et al., 2002), which then further manifests the math-gender stereotype (e.g., Jussim et al., 2005; Nosek & Smyth, 2011).

This self-perpetuating nature of the math-gender stereotype also contributes to its prevalence in educational settings (Cvencek et al., 2011; 2012; Eccles, 2011; Eagly & Koenig, 2021). While, according to some studies (e.g., Passolunghi et al., 2014), the math-gender stereotype seems to have decreased in young students, other research suggests that already primary school children associate mathematics stronger with *male* than *female* (Cvencek et al., 2011; Miller et al., 2015). Among adolescent students, this connotation of mathematics as male-typed is more robustly found across different studies (e.g., Starr et al., 2023; Starr & Simpkins, 2021; Steffens & Jelenec, 2011). This prevalence pattern suggests that the math-gender stereotype is acquired over time due to cumulative exposure to it in society, for example, by important socializers such as teachers (Eccles, 2011; Eliot, 2010; Ellemers, 2018; Gunderson et al., 2012).

Like students, teachers also hold the math-gender stereotype (Gunderson et al., 2012). Teachers express their math-gender stereotype via more positive attitudes about male students' mathematical performance, higher expectations regarding male students' success in mathematical fields, gender-biased grading in favor of male students, and by overrating male students' mathematical abilities while underrating female students' mathematical abilities (e.g., Avitzour et al., 2020; Cimpian et al., 2016; Holder & Kessels, 2017; Riegle-Crumb & Humpries, 2012; Robinson-Cimpian et al., 2014; for a literature review, see Li, 1999). Teachers' math-gender stereotype has been found to predict students' math-gender stereotype (Keller, 2001).

### **Teachers' Math-Gender Stereotype-Reinforcing Behaviors**

Teachers convey their math-gender stereotype to their students via different mechanisms. According to learning theories, gender stereotypes can be transported via observational learning (Bandura, 1977a; 1977b) or operant conditioning (Mischel, 1966; Skinner, 1963; 1977). Furthermore, according to theories about information processing, teachers' math-gender stereotype can impact student-teacher interactions in specific situations (e.g., Spencer et al., 1999).

Theories on observational learning describe that learners learn certain behaviors by observing and imitating a role model's behaviors (Bandura, 1977a). A model that is similar to the learner, for instance, because of their gender, increases the learners' *self-efficacy* (Bandura, 1977b). Self-efficacy describes a learner's personal beliefs about their capabilities in relation to the modeled behavior and, therefore, predicts the learners' motivational and self-regulatory activities to potentially imitate the modeled behavior (Bandura, 1997; Pajares, 1997). A female learner observing a female role model might think: "If she can do it, I can do it, too." Accordingly, while still underrepresented, female role models in mathematics and STEM teaching benefit female learners' self-efficacy in these areas (e.g., Dasgupta, 2011; Young et al., 2013). The stereotypical math teacher, however, is still imagined as male (Gjøvik et al., 2023), and female protagonists, also functioning as potential role models, are underrepresented in math learning materials (Jehle et al., 2024; Kerkhoven et al., 2016; Moser & Hannover, 2014; Nandi et al., 2024).

Theories on operant conditioning assume that teachers, among other socializers, reinforce learners for behaving according to the stereotypes associated with their gender (Bian, 2022; Mischel, 1966). According to operant conditioning theories, a subject is more likely to exhibit a behavior if said behavior is reinforced by a positive response to the behavior, a *reinforcement* (Skinner, 1937; 1963; 1977). In one study, for example, math teachers were found to praise female learners more frequently for their neatness than male learners while praising male learners more frequently for their mathematical performance (Corpus & Lepper, 2007). Through the lens of operant conditioning, if the girls experience this praise as reinforcement, they will more likely focus their efforts on their neatness instead of their mathematical performance since they can expect further reinforcement for the first but not for the latter. Male learners will more likely show a contrasting pattern.

Additionally, based on the assumption that teachers and their students hold similar stereotypes due to similar socialization (Martin & Halverson, 1981; Vinacke, 1957), their mutual math-gender stereotype can be activated automatically based on situational characteristics (Hilton & von Hippel, 1996). If a teacher activates it by making gender differences in mathematics obvious, for instance, via an implicit difference in praise or an explicit comment, such activation may lead to stereotype-conforming behavior like a *self-fulfilling prophecy* (Hilton & von Hippel, 1996; Jamil et al., 2018), or a *stereotype threat* (Galdi et al., 2014; Spencer et al., 1999). This means that the activated stereotype can influence the stereotyped person's behavior or performance in the situation. In terms of a stereotype threat, activation of the math-gender stereotype in a math classroom might occupy necessary cognitive resources for solving a math problem in present girls, contributing to their lowered math performance (Galdi et al., 2014). Consequences of stereotype threat may be affective and motivational (Seo & Lee, 2021; Thoman et al., 2013). Affectively, stereotype threat can contribute to female learners' *math anxiety* (Seo & Lee, 2021), meaning their elevated physiological reaction when confronted with math-related tasks and a stimulus-learned fear of mathematics (Faust, 1992). Motivationally, female learners may adopt the perspective of having weaker mathematical abilities and consider themselves less likely to succeed in mathematics, which might lead to less effort due to a perceived lack of payoff, meaning lowered success expectancy (McFarland et al., 2003; Thoman et al., 2013). Female learners' lowered success expectancy then contributes to their lowered motivation for mathematics (Eccles, 2011; Eccles & Wigfield, 2024; Wigfield & Eccles, 2000).

Altogether, teachers' math-gender stereotype and their behaviors that potentially reinforce said math-gender stereotype have short- and long-term interacting impacts on female learners' lowered interest, motivation, and performance in mathematics, which ultimately contribute to low female representation in mathematics and math-intensive STEM fields (Cvencek et al., 2012; Eccles, 2011; Gunderson et al., 2012; Halpern et al., 2007; Hand et al., 2017; Kahn & Ginther, 2017; Nosek & Smyth, 2011).

It is, therefore, important to reduce the math-gender stereotype among teachers as important socializers (Eccles, 2011; Gunderson et al., 2012; Wentzel, 2014). However, such a reduction of the math-gender stereotype has rarely been conducted successfully and sustainably (e.g., FitzGerald et al., 2019; Kollmayer et al., 2020). That might be, because stereotypes are not only useful, reducing our environments' complexity (Fiske, 1982), but they are also consolidated knowledge and beliefs (Hilton & von Hippel, 1996; Hinton, 1971), partly reflecting social realities (Jussim et al., 2005). One approach that has yet only been researched scarcely is targeting and changing aspects of the math-gender stereotype that do not reflect social reality, but that can be 'disproven.' In this vein, the math-gender stereotype could be targeted by identifying its underlying specific incorrect conceptions that can be addressed and corrected (Eitel et al., 2021), meaning *math-gender misconceptions*.

### **Math-Gender Misconceptions**

Since the actual female underrepresentation in mathematics and math-intensive STEM careers (e.g., Kahn & Ginther, 2017) contributes to the math-gender stereotype (Jussim et al., 2005; Miller et al., 2015), it can be perceived as the ‘correct’ component within the math-gender stereotype. However, the math-gender stereotype as a set of beliefs and knowledge conceptions (e.g., Ashmore & DelBoca, 1979) may also contain scientifically unsupported, ‘incorrect’ knowledge, that is, *math-gender misconceptions*. Misconceptions are highly plausible knowledge conceptions that contradict empiric evidence (Hughes et al., 2013) or established theoretical frameworks (Eitel et al., 2021). Math-gender misconceptions may, for example, provide *biologistic* explanations for the math-gender gap, meaning explanations based on biological determinism. They, for instance, assume lower *innate* mathematical talent in girls due to lower prenatal testosterone exposure in female fetuses (Baron-Cohen, 2004). Such assumptions ignore the abundant evidence *against* innate talent differences between the genders (e.g., Kersey et al., 2019; Lachance & Mazzocco, 2006; Lindberg et al., 2010) and neglect the well-established and well-researched theoretical frameworks about the influence of the math-gender stereotype on the math-gender gap (e.g., Cvencek et al., 2012; Eccles, 2011; Greenwald et al., 2002). Providing such ‘simple’ biologistic explanations for the math-gender gap while it, as aforementioned, results from complex interactions (e.g., Avolio et al., 2020; Wang & Degol, 2017) makes math-gender misconceptions appear highly plausible. Biologistic explanations for the math-gender gap furthermore suggest that the gap can and should not be narrowed since this would be ‘against human nature.’ Like this, biologistic explanations alleviate feelings of societal responsibility. Consequently, math-gender misconceptions contribute to the manifestation of the math-gender stereotype (Girelli, 2023; Rubel et al., 2024).

Since schema theories assume that all information related to a social group, such as gender, is stored within the same cognitive schema (e.g., Martin & Halverson, 1981), math-gender misconceptions can be assumed to be stored close to the math-gender stereotype. Furthermore, theories about the nature of knowledge and learning assume knowledge conceptions, like math-gender misconceptions, to be *specific* mental elements (von Aufschnaiter & Rogge, 2015). Some mental elements are considered smaller than others. Those smaller elements are organized in networks forming part of larger elements (von Aufschnaiter & Rogge, 2015). Thus, math-gender misconceptions might be smaller mental elements organized in networks within the larger mental element of the math-gender stereotype. Stereotypes, as the cognitive component of an attitude, contain knowledge and beliefs about social groups (Fiske, 1982), such as math-gender misconceptions. Accordingly, misconceptions about a topic are related to negative attitudes about the topic they refer to (Sinatra & Seyranian, 2015), highlighting the close relationship between knowledge and attitudes. Targeting and potentially refuting math-gender misconceptions might, therefore, also influence related attitudes (Jacobson et al., 2021; Thacker et al., 2020), for example, the math-gender stereotype or more specific attitudes about gender equity in mathematics. These assumed relations might make math-gender misconceptions a good starting point for learning and potential *conceptual change* (Larkin, 2012).

### **Conceptual Change Using Refutation Texts**

Conceptual change describes the process of targeting, refuting, and potentially changing a misconception (von Aufschnaiter & Rogge, 2015). Conceptual change is not trivial because knowledge conceptions are stored in long-term memory and intertwined with other mental elements (von Aufschnaiter & Rogge, 2015; Vosniadou & Skopeliti, 2014), such as knowledge conceptions (e.g., Sweller et al., 1998), attitudes, values and beliefs (e.g., Sinatra & Seyranian, 2015; Vosniadou & Mason, 2012).

In this vein, certain requirements must be met for successful conceptual change. According to Kendeou and O'Brien (2014) and their *Knowledge Revision Components Framework (KReC)*, the *co-activation* of the targeted misconception and the alternative scientifically supported explanation is necessary. Furthermore, the *integration* of the supported conception in the relevant cognitive schema must occur. Lastly, Kendeou and O'Brien (2014) describe *competing activation* as vital, meaning the misconception remains within the cognitive schema and the scientifically supported information is activated simultaneously – iteratively stronger than the misconception, while the misconception is inhibited in relevant situations. Vosniadou (1994) furthermore highlights the importance of *metacognition* for successful conceptual change: An own misconception must be recognized to be able to change it. Similarly, Strike and Posner (1982) argue that *cognitive conflict* is necessary to achieve successful conceptual change. Dissatisfaction with a prior (mis-)conception contrasting with a new, alternative conception can spark this conflict if the new conception is intelligible, plausible, and helpful. Like this, Strike and Posner (1982) touch on motivational aspects of conceptual change since (a) dissatisfaction with a prior misconception or (b) helpfulness of the new conception should foster motivation by (a) decreasing the subjective task value of applying the prior, unsupported conception or (b) increasing the subjective task value of applying the alternative, supported conception instead (see also Wigfield & Eccles, 2000).

Gregoire (2003), in her *Cognitive Affective Model of Conceptual Change (CAMCC)*, includes motivational aspects more explicitly by integrating conceptual change theories with dual process theories. Gregoire argues that several steps are necessary for successful conceptual change. The first step is *self-implication*; learners must realize they have a misconception (Gregoire, 2003). This feeling of self-implication goes hand in hand with increased stress (Gregoire, 2003). If learners appraise this stress as a challenge rather than a threat, said appraisal should increase motivation for *systematic information processing* (Gregoire, 2003; see also Maheswaran & Chaiken, 1991). Systematic information processing, according to dual process theories, involves effortful critical thinking and the (re-)construction of knowledge (Chaiken & Ledgerwood, 2012; Gregoire, 2003; Strack & Deutsch, 2015). Gregoire (2003) postulates that systematic information processing of the refutational message is necessary for conceptual change. If the information is processed predominantly *heuristically*, this hinders conceptual change (Gregoire, 2003). Heuristic information processing, according to dual process theories, is based on 'cognitive shortcuts', drawing on generalized or simplified prior knowledge (Chaiken & Ledgerwood, 2012). Other than heuristic information processing, systematic information processing must be applied intentionally (Chaiken & Ledgerwood, 2012; Strack & Deutsch, 2015), which makes motivation an essential prerequisite for systematic processing and conceptual change (Chaiken & Ledgerwood, 2012; Eagly & Chaiken, 1993; Gregoire, 2003).

*Refutation texts*, as a text form and intervention, hold the potential to fulfill the aforementioned requirements for conceptual change. Refutation texts name a commonly held misconception, state that it is not supported by evidence, and subsequently provide an alternative explanation (Tippett, 2010). Naming commonly held misconceptions can function as a metacognitive prompt, fostering metacognition by making learners reflect if they hold the misconception (Vosniadou, 1994). Furthermore, it can lead to dissatisfaction with the misconception (Strike & Posner, 1982) by providing the opportunity to detect the threshold between one's misconception, if prevalent, and the scientifically supported conception. Perceiving the potential reduction of this threshold as a *challenge* that can be mastered due to an intelligible alternative explanation can further foster motivation for systematic information processing, which is necessary for conceptual change (Gregoire, 2003; Strike & Posner, 1982). Naming the misconception followed by the scientifically supported, alternative explanation can furthermore lead to co-activation (Kendeou & O'Brien, 2014), which should help the integration of the scientifically supported, alternative explanation

within the relevant cognitive framework (Kendeou & O'Brien, 2014). The integration within the relevant cognitive framework should then foster competing activation in the long run, meaning the activation of the supported conception should become iteratively stronger while inhibiting the misconception. This process is facilitated if the alternative explanation is plausible, helpful, and applicable in a wide variety of contexts (Kendeou & O'Brien, 2014; Strike & Posner, 1982). Accordingly, compared to other text forms, such as expository texts, which explain a topic without considering commonly related misconceptions, refutation texts are a powerful tool for successful conceptual change (for meta-analyses, see Danielson et al., 2024; Tippett et al., 2010). However, most research on conceptual change so far has focused on children and younger learners (Tippett, 2010), and only recently, more research on other target groups, for example, teachers, has been introduced (e.g., Dersch et al., 2022a; Ferrero et al., 2020; see also Danielson, 2024), with some theoretical implications and empirical evidence pointing toward specific challenges regarding *teachers'* conceptual change (Dersch et al., 2022a; Gregoire, 2003).

### **Challenges for Teachers' Conceptual Change**

Teachers' knowledge and misconceptions about education, teaching, and learning are consolidated not only based on their work experience as teachers but also on their own school and university education (Gregoire, 2003), making such misconceptions appear highly plausible and therefore more robust to change (Strike & Posner, 1982). Teachers base most of their beliefs about learning and teaching on anecdotal evidence (Ferguson & Bråten, 2022; Menz et al., 2021), which has been found to predict their misconceptions (Menz et al., 2021). Moreover, teachers' certainty of their educational knowledge and potential misconceptions about teaching and learning increases with age and experience, only decreasing again for senior teachers (Munthe, 2001).

Additionally, teachers often only have limited time (Creagh et al., 2023). While student teachers' educational psychology misconceptions can be targeted in refutation lectures (Menz et al., 2024), such lectures are likely unfeasible for *in-service teachers*. A potential solution for changing in-service teachers' educational psychology misconceptions is online interventions (Dersch et al., 2022a).

Online interventions, however, occur in distracting contexts requiring high self-control (Endres et al., 2025). They may also be affected by the *screen inferiority effect* (Salmeron et al., 2024). This effect describes that information presented on a screen, like an online intervention, is not learned as well as analogous information (Salmeron et al., 2024) due to lower invested mental effort (Sidi et al., 2016). Since systematic information processing involves the *active*, critical evaluation of information (Chaiken & Ledgerwood, 2012; Gregoire, 2003), it requires active mental effort (Klepsch & Seufert, 2021). Less active mental effort due to an online intervention and, owing to that, less systematic information processing might lead to potentially inferior or even unsuccessful conceptual change.

As aforementioned, self-implication can foster systematic information processing. Teachers as experienced practitioners with limited time may, however, not feel the necessary self-implication when confronted with misconceptions about teaching and learning (Creagh et al., 2023; Dersch et al., 2022a). While the potential metacognitive cue "Many people believe that [...]. However, that is not supported by evidence." can foster systematic information processing (Gregoire, 2003), teachers might think "Many people hold this misconception, but not me" (Dersch et al., 2022a). Hence, reading refutation texts in an online intervention may not necessarily increase teachers' self-implication enough to foster their conceptual change (Dersch et al., 2022a). To address these issues of time scarcity, potential screen inferiority effects, and insufficient self-implication, an online intervention can exploit the opportunities of digital media, for instance, by applying personalized feedback based on prior performance combined with a refutation text; *personalized refutation text* (Dersch et al., 2022a).

### **Personalized Refutation Text**

*Personalized* messages can lead to higher personal involvement and more focused, systematic information processing (Mayer, 2014). Personalized *feedback* has the potential to improve learning, especially if adapted to a learner's prior performance (Narciss et al., 2014). Most research on personalized *adaptive* feedback focuses on learners' performance in a prior knowledge test (Maier & Klotz, 2022). Regarding conceptual change, such personalized feedback can highlight the threshold between a learner's individual misconception and an alternative correct conception (Asterhan & Dotan, 2018; Sánchez et al., 2009). Highlighting this threshold with personalized feedback can function as a metacognitive prompt similar to a refutation text (Vosniadou, 1994).

On the one hand, combining personalized feedback on prior performance with an expository text might be sufficient for teachers' conceptual change. Providing personalized feedback and an expository text could further be perceived as less prescriptive and authoritative than a standard refutation text (Zengilowski et al., 2021; 2022).

On the other hand, *personalized refutation text* has the potential to combine the strengths of both personalized feedback and refutation text (Asterhan & Dotan, 2018; Dersch et al., 2022a). By considering a learner's individual prerequisites, personalized refutation texts could still be less authoritative and prescriptive (Asterhan & Dotan, 2018; Zengilowski et al., 2021; 2022). Personalized feedback can make the threshold between a learner's *own misconception* and alternative, correct conception more obvious if combined with a refutation text rather than an expository text because, different from personalized expository text, personalized refutation text clearly states and disproves a misconception. Like this, personalized refutation text can increase the learner's experienced dissatisfaction with their misconception, fostering motivation for conceptual change (Strike & Posner, 1982). Providing *personalized* feedback before presenting a refutation text can also increase a learner's self-implication (Dersch et al., 2022a). Such self-implication can be expressed by negative affect of internal blame (Dersch et al., 2022a). The learner is aware of holding a misconception and feels in charge of changing it. Such internal blame can lead to more systematic information processing of the refutational message, which may drive conceptual change (Dersch et al., 2022a; Gregoire, 2003). Such increased self-implication could be especially helpful for teachers' conceptual change. Accordingly, in a study with in-service teachers, standard refutation text was as effective as expository text. Only if the refutation text was combined with personalized feedback on teachers' misconceptions, their conceptual change was improved. A slight increase in negative affect of internal blame, like guilt, mediated the effect (Dersch et al., 2022a).

The referenced study researched the refutation of multimedia learning misconceptions. The refutation text paradigm has been criticized for re-using the same refutation texts and topics, making it difficult to generalize the refutation text effect (Zengilowski et al., 2021). This criticism applies even more to *personalized* refutation text. Hence, in this dissertation, I applied personalized refutation text to another topic: math-gender misconceptions. However, targeting *math-gender* misconceptions, especially with personalized refutation text, might pose particular challenges.

### **Pitfalls and Potentials of Targeting Math-Gender Misconceptions**

Potential challenges when targeting math-gender misconceptions are based on their position and function in long-term memory. As described above, math-gender misconceptions are likely related to the math-gender stereotype (Martin & Halverson, 1981). Targeting math-gender misconceptions could, therefore, activate the math-gender stereotype. While co-activating the refutational information and the misconception itself might benefit conceptual change (Kendeou & O'Brien, 2014), simultaneous activation of the broad math-gender stereotype-related knowledge and beliefs, as well as the mere association of *mathematics* and *male*, potentially consolidated from early childhood, could outweigh or overshadow the refutational information. Drawing on this math-gender stereotype-based

oversimplified or overgeneralized knowledge and beliefs as a ‘cognitive shortcut’ is *heuristic processing* (Bodenhausen, 1990). According to the CAMCC, such heuristic processing hinders the sufficient systematic processing of the refutational message necessary for conceptual change (Gregoire, 2003).

Moreover, math-gender misconceptions combine information about mathematics and gender. This knowledge is likely related to learners’ self-concepts. The self-concept combines all knowledge schemata about the self (Sherman & Cohen, 2002). Knowledge about one’s mathematical competence, and perceived difficulty of mathematics form part of the *academic self-concept*, meaning implicit and explicit schemata about one’s academic abilities (Arens et al., 2021; Brookover et al., 1964; Ommundsen et al., 2007). As an important social characteristic and a crucial factor for socialization in most societies, gender also forms part of the self-concept (Crain, 1996; Ellemers, 2018; Stockard, 2006). Moreover, the math-gender stereotype and math-gender misconceptions may have impacted many learners’ academic pathways and careers, shaping their identities (e.g., Chow & Salmela-Aro, 2011; Eccles, 2011; Eccles et al., 1999; Girelli, 2023; Nosek & Symth, 2011; Song et al., 2017; Wang & Degol, 2017; Watt et al., 2012). Targeting their math-gender misconceptions could hence threaten learners’ identities or self-concepts.

If targeting a misconception threatens a learner’s self-concept or identity, this can provoke *worldview backfire effects* (Cook & Lewandowsky, 2011; Jacobson et al., 2021; Lewandowsky et al., 2020; Sherman & Cohen, 2002). Backfire effects hinder conceptual change or even reinforce the targeted misconception (Cook & Lewandowsky, 2011). Worldview backfire, as one kind of backfire effect, is a learner’s defensive reaction to a *threat* to their *identity* or *self-concept* (Lewandowsky et al., 2020; Sherman & Cohen, 2002).

Worldview backfire hindering successful conceptual can take effect via two different mechanisms. First, learners may react with repulse or discount the refutational information (Chinn & Brewer, 1993; Gregoire, 2003). This reaction leads to a lack of motivation to process the refutational information systematically, which would be necessary for conceptual change (Gregoire, 2003). As a consequence, learners process the refutational information heuristically, which is insufficient for conceptual change (Gregoire, 2003). Hence, repulse hinders conceptual change (Chinn & Brewer, 1993). This mechanism *protects* a learner’s self (Chinn & Brewer, 1993).

The second mechanism is the *disconfirmation bias* (Edwards & Smith, 1996). If a disconfirmation bias occurs, the learners process the refutational information systematically. However, in doing so, they actively argue against the refutational information. Thus, the disconfirmation bias strengthens the targeted misconception (Cook & Lewandowsky, 2011). This mechanism *defends* a learner’s self (Edwards & Smith, 1996; Sherman & Cohen, 2002). Since both mechanisms might interact or affect only parts of a refutational message, backfire effects might lead to reduced conceptual change, no conceptual change, or the reinforcement of targeted misconceptions.

The math-gender stereotype is the cognitive component of an attitude and is therefore related to its affective component (Fiske, 1982). Consequently, activating the math-gender stereotype by targeting math-gender misconceptions may also provoke an affective reaction. If a refutational message is perceived as a threat to identity, this likely provokes negative affect of *external blame*, like anger (Trevors et al., 2016). More anger is related to stronger backfire effects (Trevors et al., 2016). Furthermore, backfire effects may not only reinforce the misconceptions themselves, like the targeted math-gender misconceptions, but also the related attitudes (Jacobson et al., 2021), for instance, the math-gender stereotype or specific attitudes towards gender equity in mathematics. *Knowledge backfire* describes the ironic reinforcement of a misconception after an attempted refutation, whereas *attitude backfire* describes the ironic reinforcement of a related attitude (Jacobson et al., 2021).

According to the CAMCC (Gregoire, 2003), combining personalized feedback with refutation text can enhance conceptual change due to increased self-implication, which causes stress. However, this is only the case if the stress is

appraised as a challenge rather than a threat (Gregoire, 2003). If the refutational information is inconsistent with a learner's self-concept or identity, the stress caused by self-implication is likely appraised as a threat (Sherman & Cohen, 2002). Such a threat appraisal can hinder conceptual change or trigger backfire effects (Cook & Lewandowsky, 2011; Gregoire, 2003). Potential worldview backfire effects due to targeting math-gender misconceptions might, therefore, be even stronger for *personalized* refutation text (Dersch et al., 2022a). In sum, aiming to refute *math-gender misconceptions* might endanger conceptual change or lead to stronger worldview backfire effects than refuting misconceptions on a topic less related to identity and self-concept. This might be even more so when applying *personalized* refutation texts.

Nevertheless, addressing these math-gender misconceptions is important since they potentially contribute to the manifestation of the math-gender stereotype, contributing to female underrepresentation in mathematics (e.g., Girelli, 2023). Targeting and refuting math-gender misconceptions might change related detrimental attitudes (see also Jacobson et al., 2021; Thacker et al., 2020), such as the math-gender stereotype or more specific attitudes towards gender equity in mathematics. Therefore, while addressing math-gender misconceptions with personalized refutation texts holds pitfalls, it also holds potentials for female representation in mathematics.

### **Goals of the Present Dissertation**

Math-gender misconceptions are scientifically unsupported but highly plausible 'explanations' contributing to the manifestation of the math-gender stereotype and the math-gender gap (see also Girelli, 2023). Thus, they should be targeted and refuted among teachers as influential socializers. To do so, it is important to know and measure them specifically (Eitel et al., 2021). The first goal of this dissertation was to identify math-gender misconceptions, to research their prevalence among teachers, and to explore their relation to the math-gender stereotype.

Moreover, I was interested in the relationship between math-gender misconceptions and behaviors. Since teachers' misconceptions about teaching and learning can influence their behaviors (e.g., McAfee & Hofmann, 2021; Wilen, 2004), *math-gender* misconceptions might too. Teachers could, informed by their math-gender misconceptions, reinforce the math-gender stereotype among their students through explicit statements but also through subtle, implicit behaviors (Bian, 1922; Beilock et al., 2010; Eccles, 2011; Gunderson et al., 2012; Mischel, 1966). Since researching a broad range of math-gender stereotype-reinforcing behaviors in the field is challenging, I researched teachers' *awareness* of such behaviors. Awareness of math-gender stereotype-reinforcing behaviors can help teachers critically reflect on and potentially modify their own behaviors. I also investigated the relationship between math-gender misconceptions and math-gender stereotype-reinforcing behaviors.

Another goal of this dissertation was to study if and how teachers' math-gender misconceptions can be targeted and potentially changed by combining refutation texts with personalized feedback. In doing so, I researched how refutation materials can be enhanced for the target group of teachers (Gregoire, 2003). Apart from researching the interventions' influence on teachers' math-gender misconceptions (i.e., knowledge), I also researched its influence on teachers' awareness of math-gender stereotype-reinforcing behaviors (i.e., application of knowledge; Duit & Treagust, 2003) and the potential influence of conceptual change on teachers' related attitudes, like their math-gender stereotype and their specific attitudes towards gender equity in mathematics (Cvencek et al., 2012; Martin & Halverson, 1981; Nosek & Smyth, 2011). However, since math-gender misconceptions are potentially related to teachers' identity and self-concept, targeting them with a refutation instruction might trigger backfire effects (e.g., Edward & Smith, 1996; Jacobson et al., 2021; Sherman & Cohen, 2002). In this dissertation, I therefore also accounted for potential knowledge and attitude backfire effects.

## Summaries of Studies Included in this Dissertation<sup>1</sup>

### Study 1

**Dersch, A.-S., Heyder, A., & Eitel, A. (2022b). Exploring the Nature of Teachers' Math-Gender Stereotypes: The Math-Gender Misconception Questionnaire. *Frontiers in Psychology, 1728*.**  
<https://doi.org/10.3389/fpsyg.2022.820254>

#### *Theoretical Background and Study Motivation*

In the first study of this dissertation, we created and tested a questionnaire for math-gender misconceptions. Based on the literature, we identified three repeatedly appearing, empirically unsupported conceptions potentially associated with the math-gender stereotype. All three misconceptions try to 'explain' gender differences in mathematical talent, ignoring the evidence for talent similarities rather than differences between the genders (e.g., Kersey et al., 2019; Lachance & Mazzocco, 2006).

The first misconception is *empathizing-systemizing*. It attributes assumed math talent differences between the genders to higher pre-natal testosterone exposure in male fetuses (e.g., Baron-Cohen, 2004). Such testosterone exposure is believed to increase systematic thinking abilities and simultaneously reduce empathetic thinking abilities in male fetuses. Lower testosterone, in contrast, is believed to cause a contrasting pattern in female fetuses (Baron-Cohen, 2004). This misconception, hence, proposes that empathetic and systematic thinking are dualities. However, both are related to general intelligence (Gottfredson, 1998; Martingano & Konrath, 2022). Furthermore, this misconception proposes that systematic thinking abilities, which girls are assumed to lack, predict mathematical talent (e.g., Baron-Cohen, 2004). This relation is, however, disproven (Escovar et al., 2016).

The second misconception is *girls' compensation*. This misconception assumes that girls work harder than boys for similar achievements in mathematics, having to compensate for their lower innate talent. While girls work harder than boys in mathematics, this leads to *better-* not similar grades (Duckworth & Seligman, 2006). Besides, a need for compensation is inconclusive (Girelli, 2023; Kersey et al., 2019; Lachance & Mazzocco, 2006).

The third misconception is *girls' non-compensability*. It is related to beliefs about talent not being malleable. Such a *fixed mindset* is broadly prevalent in mathematics (e.g., Canning et al., 2019; Gunderson et al., 2017; Heyder et al., 2021; Leslie et al., 2015). A fixed mindset is at odds with evidence of educational achievement, such as the *growth mindset*, which proposes that skills can be improved with practice (Dweck, 1999; 2015). The *girls' non-compensability* misconception combines the idea of a fixed mindset with attributing girls less mathematical talent. In this way, girls are assumed to be unable to compensate for mathematical talent differences. The combination of these assumptions is detrimental to female students' attitudes related to mathematics, to their motivation, and, in consequence, also to their mathematical performance (Canning et al., 2022; Heyder et al., 2019, 2021; Muenks et al., 2020).

Those three misconceptions have a substantial theoretical overlap but still represent different aspects. According to theories about knowledge and stereotypes, these math-gender misconceptions are likely related to the math-gender stereotype (Martin & Halverson, 1981; von Aufschnaiter & Rogge, 2015).

#### *Hypotheses*

We investigated to what degree these three potential misconceptions are empirically separable, assuming three independent but related math-gender misconceptions to cluster in three misconception factors (*empathizing-systemizing*, *girls' compensation*, and *girls' non-compensability*). We were also interested in their prevalence. Furthermore, we researched the relation of these math-gender misconceptions to the math-gender stereotype.

---

<sup>1</sup> Analyses in all studies within this dissertation were conducted with an  $\alpha$ -level of .05. We reported effect sizes for correlations, contrast analyses, and ANOVAs based on Cohen (2013).

### **Method**

We developed the *Math-Gender Misconception Questionnaire (MGMQ)* and empirically evaluated it in an online survey with 303 pre-service teachers (242 cisgender female, 61 cisgender male,  $M_{\text{age}} = 21.73$ ). The MGMQ consisted of 30 items, of which 50 % (15) were misconception items for the scales of *empathizing-systemizing* (five items, e.g., “As girls think rather empathetically and boys think rather systematically, boys are on average more talented in mathematics than girls.”), *girls’ compensation* (five items, e.g., “For the same grades in mathematics, boys, on average, work less hard than girls, as boys have more natural talent.”), and *girls’ non-compensability* (five items, e.g., “With their, on average, harder work, girls can still not fully compensate their lesser mathematical talent compared to boys.”). The correct answer was to disagree with them. We balanced the questionnaire with 50 % (15) items comprising correct statements (e.g., “On average, boys rate themselves as more competent than girls in terms of their mathematical abilities.”). We did not analyze the filler items since they did not provide information about the prevalence of specific misconceptions. Each item required verification of a statement (*agree/disagree*) and certainty regarding the answer (0 = *very uncertain* to 4 = *very certain*). We multiplied verification (-1 = disagreement to a misconception item; +1 agreement to a misconception item) and certainty (0–4), resulting in a misconception score (-4 = no misconceptions, 0 = uncertainty, +4 = strong misconceptions; cf. Eitel et al., 2021). We also measured misconception prevalence with a cutoff criterion: If at least one misconception item per scale was agreed to with high certainty (3–4), we assumed at least a (mixed) misconception to be prevalent (see also Dersch et al., 2022a; Vosniadou, 1994). We also analyzed the correlation of the researched misconceptions with the math-gender stereotype measured with one global item (“*Mathematics is rather...*” on a 9-point Likert-scale from *female* to *male*).

### **Results and Discussion**

A confirmatory factor analysis revealed a good fit for the three-factor model,  $CFI = .940$ ,  $RMSEA = .058$ ,  $SRMR = .057$  (Hu & Bentler, 1999; Yu, 2002). All factor loadings were significant ( $ps < .01$ ) and ranged between .44 and .83. Scale reliabilities were in a good-to-acceptable range for *empathizing-systemizing* ( $\omega = .88$ ), *girls’ compensation* ( $\omega = .76$ ), and *girls’ non-compensability* ( $\omega = .72$ ; Nunnally, 1978). Girls’ compensation and girls’ non-compensability were strongly correlated ( $r = .86$ ,  $p < .001$ ). Misconception scores were negative (*empathizing-systemizing*:  $M = -0.70$ ,  $SD = 2.44$ ; *girls’ compensation*:  $M = -1.58$ ,  $SD = 2.15$ ; *girls’ non-compensability*:  $M = -2.34$ ,  $SD = 1.90$ ), the majority of participants did not hold strong misconceptions. However, 48.2 % of pre-service teachers endorsed at least one of three misconceptions according to the cutoff criterion. Specifically, 32.0 % believed boys are inherently better in mathematics because they think more systematically (*empathizing-systemizing*), 26.7% believed that girls are only as good in mathematics as boys because they work harder (*girls’ compensation*), and 17.5% believed that girls cannot compensate for their lower mathematical abilities (*girls’ non-compensability*). The correlation between math-gender misconceptions and the math-gender stereotype was moderate,  $r = .45$ ,  $p < .001$ , suggesting that both are related but not the same.

In sum, the results of this first study supported the idea of three misconceptions related to the math-gender stereotype prevalent among a relevant subsample of pre-service teachers. Misconception prevalence called for an intervention targeting math-gender misconceptions, for example, in teacher training. An intervention would be especially useful and relevant if math-gender misconceptions informed or motivated teachers’ behaviors. Whether this was the case remained unclear after this first study. Therefore, it was necessary to study teachers’ math-gender stereotype-reinforcing behaviors or, more specifically, awareness of such behaviors as a potential close-to-behavioral measure.

## Study 2

### Dersch, A.-S., Heyder, A., & Eitel, A. (submitted). **Teachers' Awareness of Math-Gender Stereotype-Reinforcing Behaviors: A Video Experiment**

[currently under review at *Social Psychology of Education*]

#### ***Theoretical Background and Study Motivation***

In this study, we researched a close-to-behavioral measure potentially related to math-gender misconceptions. That is *awareness of math-gender stereotype-reinforcing behaviors*. Teachers are important socializers contributing to their students' math-gender stereotype (e.g., Eccles et al., 2011; Gunderson et al., 2012; Keller, 2001). Teachers' implicit math-gender stereotype-reinforcing behaviors include the selection of stereotypical math tasks, which is likely since math textbooks depict less women than men and describe men more often via their STEM profession than women (e.g., Jehle et al., 2024; Nandi et al., 2024; Moser & Hannover, 2014). Another example of implicit math-gender stereotype-reinforcing behaviors is differences in praise between the genders (e.g., Corpus & Lepper, 2007; Gunderson et al., 2012; Jamil et al., 2018). More explicit math-gender stereotype-reinforcing behaviors would be overt teacher comments encouraging only boys or stating that comparing the genders in math is unfair, indicating a fixed mindset in mathematics, which is detrimental for female learners (e.g., Canning et al., 2022). Teachers should be aware of such behaviors to avoid expressing them and to detect and critically evaluate them in others.

Teachers' detection and critical evaluation of relevant behaviors are related to their professional knowledge (Kersting et al., 2012). Videos are helpful for researching teachers' detection and critical evaluation of relevant teacher behaviors (Bergin et al., 2017; Chan et al., 2021). Hence, *awareness*, as a combination of critical evaluation and detection of relevant *math-gender stereotype-reinforcing behaviors* in other teachers, might complement a misconception questionnaire like the MGMQ.

Apart from researching teachers' awareness of math-gender stereotype-reinforcing behaviors and the relation of awareness to the math-gender stereotype and math-gender misconceptions, we were interested in other social factors potentially influencing awareness. Awareness of gender stereotypes is lower in people socialized as male than in people socialized as female (Becker & Swim, 2011), which might be similar to the *math-gender stereotype*. Furthermore, knowledge of the math-gender stereotype being harmful to female learners (e.g., Eccles, 2011) might lead to less awareness of a female model teacher's math-gender stereotype-reinforcing behaviors. Lastly, the model teacher's gender and the participating teachers' gender being similar might improve awareness of math-gender stereotype-reinforcing behaviors since model-observer-similarity increases a learner's attention to a model (Bandura, 1977a).

#### ***Hypotheses***

We researched participating teachers' awareness of math-gender stereotype-reinforcing behaviors. We expected teachers to be more aware of a male model teacher's math-gender stereotype-reinforcing behaviors than of the same behaviors in a female model teacher. We also researched if participating teachers' gender influenced their awareness, expecting awareness to be lower among male participating teachers. Additionally, we were interested in model-observer-gender similarity, meaning the interaction of model teacher gender and participant teacher gender. Moreover, we researched the relation of math-gender stereotype-reinforcing behaviors with math-gender misconceptions and the math-gender stereotype.

#### ***Method***

Participants were 289 teachers (119 pre-service teachers, 170 in-service teachers; 54 cisgender male, 223 cisgender female, two non-binary – socialized as female;  $M_{\text{age}} = 37.0$ ). They watched four scripted video sequences

(0:11 – 2:10 min). Participants were randomly assigned to the model teacher in all videos being male ( $n = 130$ ) or female ( $n = 149$ ). Apart from the model teacher's gender, the video scripts did not differ. All participants watched the sequences in the same order, starting with implicit behaviors (first two videos: selection of stereotypical math problems, differences in praise) and increasing in explicitness of the math-gender stereotype-reinforcing behaviors (last two videos: explicit encouragement for boys only, explicit statement that comparing the genders in math was unfair). The videos also contained distractor behaviors, such as the teacher using a random number generator to decide which student should answer their question. After each video, the participants evaluated the model teacher's behavior and teaching strategies in general with one open question. Two independent coders coded participants' evaluations of the model teachers' behavior regarding the mere detection of math-gender stereotype-reinforcing behaviors (0; 1) and how they critically evaluated the model teachers' behaviors and their consequences in the context of math-gender stereotypes and their societal effects (-3 to +3). We added both measures to assess participant teachers' *awareness*. Interrater reliability was good:  $\kappa_{\text{Cohen}} = .77$ ,  $z = 29$ ,  $p < .001$ . We assessed participants' math-gender misconceptions with a short version of the MGMQ, the MGMQ-s, measuring the two misconception scales of *empathizing-systemizing* and *girls' compensation*, with four items each. We also assessed participants' math-gender stereotype with the same single-item measure as in Study 1.

### **Results and Discussion**

Participating pre-service and in-service teachers detected only 1.8 – 4.7 % of *implicit* math-gender stereotype-reinforcing behaviors and 64.4 – 93.3 % of *explicit* math-gender stereotype-reinforcing behaviors. In this sample, 35.3 % of teachers held at least one math-gender misconception.

As hypothesized, participants were more aware of math-gender stereotype-reinforcing behaviors if the model teacher was male ( $M = 1.31$ ,  $SD = 0.65$ ) rather than female ( $M = 0.96$ ,  $SD = 0.60$ ),  $F(1, 272) = 23.06$ ,  $p < .001$ ,  $\eta^2 = .08$ , indicating a medium effect. Male and female participants, contrary to our hypothesis, did not differ in their awareness,  $F(1, 272) = 0.31$ ,  $p = .581$ , yielding the exact same awareness ( $M = 1.13$ ,  $SD = 0.65$ ). However, model-observer gender similarity mattered,  $F(1, 272) = 4.965$ ,  $p = .027$ ,  $\eta^2 = .02$ , indicating a small effect. We found a significant but small correlation of awareness with math-gender misconceptions,  $r = .14$ ,  $p = .017$ , making it an appropriate close-to-behavioral measure, adding value to a misconception questionnaire like the MGMQ(-s). Such a close-to-behavioral measure tests knowledge application and counters impression management based on social desirability (Kersting et al., 2012; van de Mortel, 2008).

The low awareness of implicit math-gender stereotype-reinforcing behaviors called for teacher training to raise awareness. A conceptual change intervention, for instance, could target math-gender misconceptions and potentially indirectly improve awareness of math-gender stereotype-reinforcing behaviors. It remained to be researched how such a conceptual change intervention could be designed. This was the main goal of the third dissertation study.

### **Study 3**

**Dersch, A.-S., Thacker, I., & Eitel, A. (submitted). Should We Tell Teachers That They Are (In-)correct? How Personalized Refutation Text Promotes or Hampers Conceptual Change**

[currently under review at *Educational Psychology Review*]

#### ***Theoretical Background and Study Motivation***

In the third study of this dissertation, we created an intervention to target and potentially refute teachers' math-gender misconceptions. Teachers as influential socializers were found to hold math-gender misconceptions in

the previous studies within this dissertation and said misconceptions were related to the math-gender stereotype and math-gender stereotype-reinforcing behaviors. Hence, refuting teachers' math-gender misconceptions appeared necessary. Refuting misconceptions requires systematic information processing (Gregoire, 2003). However, teachers might not process the refutational information systematically enough (Gregoire, 2003), for instance, due to not feeling self-implicated (Dersch et al., 2022a; Gregoire, 2003). Thus, standard refutation text, which is an effective means for conceptual change among novice learners (Danielson et al., 2024; Tippett, 2010), may not be sufficient for teachers' conceptual change (Dersch et al., 2022a; Ferrero et al., 2020). Combining refutation text with personalized feedback on teachers' prior performance has been found to successfully foster their conceptual change about multimedia learning (personalized refutation texts; Dersch et al., 2022a), mediated by increased self-implication, expressed via negative affect of internal blame (Dersch et al., 2022a). However, the topic of gendered ability differences might cause worldview backfire effects (Lewandowsky et al., 2020) due to endangering learners' identity or self-concept (e.g., Sherman & Cohen, 2002). Such worldview backfire effects might be related to negative affect of external blame, such as anger (Trevors et al., 2016).

### ***Hypotheses***

Accounting for the beneficial effects of personalized refutation texts for teachers' conceptual change in previous research, as well as for potential backfire effects due to the researched topic, we preregistered contrasting hypotheses. We expected personalized refutation texts would work best for teachers' conceptual change of math-gender misconceptions. We hypothesized this effect to be mediated by negative affect of internal blame (e.g., guilt). Opposed to these hypotheses, we assumed (personalized) refutation texts would cause stronger backfire effects. We assumed backfire effects to be mediated by negative affect of external blame (e.g., anger). We also researched the effects of condition on the close-to-behavioral measure of awareness of math-gender stereotype-reinforcing behaviors. Furthermore, we explored potential effects on specific attitudes towards gender equity in math and the global math-gender stereotype.

### ***Method***

In a randomized 2x2-pre-post-design with 336 teachers (85 in-service teachers, 251 pre-service teachers; 266 cisgender female, 69 cisgender male, one non-binary – socialized as female;  $M_{\text{age}} = 27.56$  [ $SD = 12.77$ , range = 18—75 years]), participants were randomly assigned to one of four text type conditions (1) standard expository texts ( $n = 83$ ), (2) personalized expository text ( $n = 86$ ), (3) standard refutation text ( $n = 74$ ), or (4) personalized refutation text ( $n = 93$ ). Participants in the two personalized text conditions received personalized feedback based on their performance in the pre-test (i.e., misconceptions as assessed in the MGMQ-s), either combined with refutation text or expository text, depending on their condition. Participants filled in the MGMQ-s as pre-test and post-test.

We researched which text type was most effective in fostering conceptual change based on the difference between the post-test and pre-test misconception scores in the MGMQ-s (max. backfire effects = -8; no conceptual change = 0; max. conceptual change = +8). After the MGMQ-s as a post-test, we also tested the effect of text type on awareness of math-gender stereotype-reinforcing behaviors in a second post-test to research knowledge application (measured as in Study 2). We then assessed participants' explicit math-gender stereotype (one-item measure, as in Study 1 and Study 2) and their specific attitudes towards gender equity in mathematics (e.g., "The math-gender gap should be closed") to research potential condition effects on attitudes and the potential beneficial effects of conceptual change on attitude change. We also investigated whether negative affect of internal blame mediated successful conceptual change. Furthermore, we tested if (personalized) refutation text triggered stronger backfire effects, potentially mediated by negative affect of external blame.

### ***Results and Discussion***

Overall, the interventions fostered teachers' conceptual change,  $t(335) = 18.72, p < .001$ . For teachers who had held misconceptions in the pre-test ( $n = 96$ ), according to orthogonal contrast analyses, the focal contrast  $[-1 -1 -1 +3]$  revealed that personalized refutation texts initiated the strongest conceptual change ( $M = 2.06, SD = 1.97$ ),  $b = .26, SE = .13, t(92) = 4.16, p = .044, R^2 = .05$ , indicating a small effect. According to the residual contrasts ( $[-1 -1 +2 0]; [-1 +1 0 0]$ ), the other conditions (standard expository text:  $M = 1.77, SD = 1.90$ ; personalized expository text:  $M = 1.80, SD = 1.53$ ; standard refutation text:  $M = 1.98, SD = 2.01$ ) did not differ significantly,  $p > .05$ . Among these 96 participants with both misconceptions in the pre-test, negative affect of internal blame did not mediate the beneficial effect of personalized refutation text on conceptual change,  $ab = -0.2, SE = .02, BCa CI 95 [-.08, .02]$ . However, researching all 336 participants, conceptual change correlated with negative affect of internal blame,  $r = .14, p = .009$ . This finding suggests a small but meaningful relationship between internal blame and conceptual change in line with previous research (Dersch et al., 2022a; Gregoire, 2003). Yet, this relationship was not due to text type. It might have been due to the topic of math-gender misconceptions increasing negative affect of internal blame.

Personalized refutation texts, while refuting teachers' misconceptions most effectively, also triggered the strongest backfire effects among all participants with negative conceptual change ( $n = 42$ ) according to an ANOVA,  $F(1, 38) = 8.11, p = .007, \eta^2 = .18$ , indicating a large effect. Most of these 42 participants had not held (strong) misconceptions in the pre-test. Negative affect of external blame did not mediate these backfire effects,  $ab = -.0002, SE = .02, BCa CI 95 [-.02, .03]$ .

Regarding the application of knowledge, the 96 participants who had held both misconceptions in the pre-test were less aware of math-gender stereotype-reinforcing behaviors if they were in the personalized refutation text condition,  $b = -.09, SE = .04, t(78) = -2.01, p = .047, R^2 = .06$ , indicating a medium effect. This finding might indicate backfire effects and is similar to results in a study by Ferrero et al. (2020), who found backfire effects on teachers' knowledge transfer in a refutation text condition, which had initially fostered the strongest conceptual change.

We did not find condition effects on the math-gender stereotype, nor on specific attitudes toward gender equity in mathematics ( $ps > .05$ ). Math-gender misconceptions in the post-test were, however, significantly negatively correlated with specific positive attitudes towards gender equity in math with a small effect,  $r = -.20, p < .001$ . The fewer misconceptions participants held after the intervention, the more positive their attitudes toward gender equity in math. Further, the math-gender stereotype did not correlate significantly with math-gender misconceptions in the post-test ( $p > .05$ ). The math-gender stereotype, thus, seemed to have remained stable. We further found that among participants with knowledge backfire effects ( $n = 42$ ), personalized feedback led to more positive specific attitudes towards gender equity in math  $F(1, 32) = 5.92, p = .021, \eta^2 = .16$ , indicating a large effect.

### **General Discussion**

In this dissertation, I aimed to construct a questionnaire that reliably measures teachers' math-gender misconceptions. I also researched math-gender misconceptions' potential relation to teachers' math-gender stereotype. Furthermore, I aimed to construct an appropriate close-to-behavioral measure related to math-gender misconceptions; teachers' awareness of math-gender stereotype-reinforcing behaviors. Additionally, this dissertation aimed at refuting teachers' math-gender misconceptions using personalized refutation texts (vs. personalized expository text, standard refutation text, and standard expository text), also testing the application of knowledge with the close-to-behavioral measure of awareness and the impact on attitudes such as the math-gender stereotype and more specific attitudes toward gender equity in mathematics. I furthermore took potential knowledge and attitude backfire effects into account.

### **Pattern of Results of this Dissertation**

Within this dissertation, I identified and measured knowledge conceptions that contradict scientific evidence and theoretical frameworks (Eitel et al., 2021; Hughes et al., 2013) about gender talent differences in mathematics, that is, math-gender misconceptions (i.e., *empathizing-systemizing*, *girls' compensation*, and *girls' non-compensability*). I found them to be related but separate constructs, which were moderately correlated with teachers' math-gender stereotype. Math-gender misconceptions were prevalent in a substantial group of pre-service and in-service teachers. Further, they were related to awareness of math-gender stereotype-reinforcing behaviors, with a small but significant effect. This finding called for an intervention refuting teachers' math-gender misconceptions, potentially increasing teachers' awareness and changing teachers' attitudes.

Therefore, I researched how these math-gender misconceptions could best be refuted and found that personalized refutation texts are a promising method, specifically for teachers with strong misconceptions. I, however, also found that personalized refutation texts unintendedly strengthened math-gender misconceptions, mostly in participants who did not hold strong math-gender misconceptions in the pre-test. Studying the application of knowledge after the intervention, I also found backfire effects for personalized refutation texts in awareness of math-gender stereotype-reinforcing behaviors for participants with both misconceptions. We did not find condition effects on attitudes like the global math-gender stereotype nor specific attitudes toward gender equity in mathematics.

### **Addressing Gaps from Previous Research**

In this paragraph, I will discuss the gaps I identified in previous research and how this dissertation attempted to target and potentially narrow these gaps. Gaps reach from methodological questions of math-gender misconception and math-gender stereotype-reinforcing behavior assessment via their prevalence and interrelation to potential effects of an intervention countering math-gender misconceptions and relations to affective processes and attitudes. I will address each gap individually by combining and synthesizing the contributions of all three studies included in this dissertation. First, I will describe the respective open question. Then, I will describe how I attempted to answer it. I will then discuss the findings and evaluate their impact in relation to the respective gap. In doing so, I will also briefly discuss research gaps that remain after this dissertation, either due to limitations of my contributions or additional questions that were beyond the scope of this dissertation.

#### ***Measuring Teachers' Math-Gender Misconceptions***

The first gap this dissertation targeted is the assessment of math-gender misconceptions. Thus far, most research aiming to research the explicit math-gender stereotype has used one global item (e.g., Nosek & Smyth, 2011; Starr et al., 2023) or two global items – one for boys and one for girls (e.g., Steffens & Jelenec, 2011), or had participants judge the same item about performance or talent differences between the genders from different perspectives (e.g., Passolunghi et al., 2014). An example item is “Mathematics is rather...” with a five-point Likert scale from *female* to *male* (Nosek & Smyth, 2011). The above-mentioned general measures include the current social reality that shapes stereotypes (Jussim et al., 2005), like the lower female representation in mathematics and math-intensive STEM fields (Kahn & Ginther, 2017; Wang & Degol., 2017). Applying such general measures appropriately represents the global nature of the math-gender stereotype (Greenwald et al., 2002; Klineberg, 1951). However, these measures do not account for specific aspects of the math-gender stereotype that are not empirically supported (Hughes et al., 2013) or not in line with theoretical frameworks (Eitel et al., 2021), meaning math-gender misconceptions. In previous research, I could not identify a measure for such misconceptions. Because misconceptions can be effectively targeted using refutation text, which first states a common misconception before

addressing it (e.g., Tippett, 2010), it is necessary to know the exact misconceptions to target them (Eitel et al., 2021). Therefore, in Study 1 of this dissertation, I constructed the Math-Gender Misconception Questionnaire (MGMQ; Dersch et al., 2022b) to assess three misconceptions I had identified within the literature. The misconceptions were *empathizing-systemizing* (e.g., “As girls think rather empathetically and boys think rather systematically, boys are more talented in mathematics than girls.”), *girls’ compensation* (e.g., “For the same grades in mathematics, boys, on average, work less hard than girls, as boys have more natural talent.”), and *girls’ non-compensability* (e.g., “With their harder work, girls can still not fully compensate their lesser mathematical talent compared to boys.”).

As expected, these misconceptions were empirically separable but related constructs (Study 1). For research with in-service teachers as a busy target group (Creagh et al., 2023), I constructed the MGMQ-s to provide a short version of the questionnaire focusing on the two more prevalent misconceptions of empathizing-systemizing and girls’ compensation (Study 2). The MGMQ-s revealed a similar prevalence of the two researched math-gender misconceptions among in-service and pre-service teachers (Study 1, Study 2, Study 3). In all three studies, I found math-gender misconceptions related to the global math-gender stereotype. These findings align with theoretical assumptions about how gender-related knowledge is stored (Martin & Halverson, 1981). My findings imply that math-gender misconceptions might contribute to the manifestation of the related math-gender stereotype, for instance, by providing simple biologicistic explanations for the math-gender gap (see also Girelli, 2023; Rubel et al., 2024).

While I identified three specific math-gender misconceptions based on the literature (e.g., Baron-Cohen, 2004; Martingano & Konrath, 2022; Sáinz et al., 2020; Tiedemann et al., 2002), which I then measured in pre-service and in-service teachers throughout all three studies, these three math-gender misconceptions may not be exhaustive. Thus, there might be other math-gender misconceptions prevalent in teachers and the general public that justify the math-gender stereotype and hence contribute to its manifestation (see also Girelli, 2023; Rubel et al., 2024), while being scientifically unsupported (Eitel et al., 2021; Hughes et al., 2013).

In sum, the dissertation resulted in a reliable questionnaire to assess frequent math-gender misconceptions, comprising both a validated long and short version of the questionnaire, the MGMQ, and the MGMQ-s. Both questionnaires allow for the assessment of specific math-gender misconceptions. However, it remained to be researched if the measured math-gender misconceptions would be related to behaviors or, more specifically, a close-to-behavioral measure, for instance, awareness of math-gender stereotype-reinforcing behaviors.

### ***Measuring Teachers’ Awareness of Math-Gender Stereotype-Reinforcing Behaviors***

Teachers’ misconceptions about learning and teaching may be related to lesson design, teacher behaviors, or student-teacher interactions (e.g., McAfee & Hofmann, 2021; Wilen, 2004). However, it remained to be researched if this holds also true for *math-gender* misconceptions. I attempted to answer this question with the help of the close-to-behavioral measure I constructed in Study 2. Since teachers can convey their stereotype to their students via different math-gender stereotype-reinforcing behaviors (e.g., Bandura, 1977a; 1977b; Bian, 2022; Beilock et al., 2010; Corpus & Lepper, 2007; Mischel, 1966; Jamil et al., 2018; Jehle et al., 2024; Skinner, 1963; 1977), I researched teachers’ awareness of math-gender stereotype-reinforcing behaviors as a potential close-to-behavioral measure.

Previous research has mostly used case vignettes to assess gender awareness in mathematics and other STEM subjects (e.g., Atanasova et al., 2024; Holder & Kessels, 2017; Nürnberger et al., 2016). These case vignettes can be re-read several times. Furthermore, short, subtle, and implicit occurrences must be described explicitly in the

vignette text, likely making implicit occurrences more obvious to the vignette readers. A potential solution to this issue could be the application of classroom videos.

In classroom video research, different protagonists show several behaviors simultaneously, some of them implicit or subtle, which allows for the ecological assessment of teacher noticing or teacher awareness of a breadth of different behaviors and interactions (Chan et al., 2021). Previous video research on teacher noticing, to my best knowledge, has focused only scarcely on awareness of a model teacher's gender stereotype-reinforcing behaviors but mainly on the perception and evaluation of student behaviors in classroom settings (Chan et al., 2021). It is, however, important for teachers to be aware of math-gender stereotype-reinforcing behaviors that could convey the math-gender stereotype to students (e.g., Bian, 2022; Beilock et al., 2010) since such awareness might be required in socialization contexts like schools (Wentzel, 2014), to detect it in colleagues, but also to reflect own behaviors.

This gap in the previous research led me to construct a measure assessing awareness of math-gender stereotype-reinforcing behaviors. More specifically, I scripted and then filmed classroom videos of a model teacher expressing different math-gender stereotype-reinforcing behaviors in a realistic math-classroom setting of a 7<sup>th</sup> grade, balanced for a male and a female model teacher. I extracted the math-gender stereotype-reinforcing behaviors for the video scripts from a literature review, researching common ways teachers might convey the math-gender stereotype to their students. I included the selection of a gender-stereotypical math task. Such tasks are still common in current math textbooks and influence female learners' perception of mathematics as male (e.g., Jehle et al., 2024; Moser & Hannover, 2014). Furthermore, I included a teacher praising girls exaggeratedly for easy math tasks while praising boys less for difficult math tasks. This behavior implies that the teacher attributes a girl's solution of an easier math task to strong effort requiring praise while attributing a boy's solution of a rather difficult math task to average effort and, hence, to more talent (see also Tiedemann, 2002). Such rather implicit or subtle behaviors likely reinforce the math-gender stereotype, for instance, in terms of self-fulfilling prophecies (Jamil et al., 2018). Two other videos contained more explicit math-gender stereotype-reinforcing behaviors, meaning direct statements. Statements differed in explicitness. The first statement occurred before an exam comparing the math performance of different classes of the same grade. The teacher stated to rely on the boys' math proficiency without addressing the girls. The second statement was about girls' and boys' math performance not being comparable due to different prerequisites of the genders (see also fixed mindset ideas; researched by Canning et al., 2022; Heyder et al., 2021).

In order to increase ecological validity, videos could only be watched once, just like real-life situations cannot be rewound. For that same purpose, all videos contained distractor behaviors; for example, the teacher used a random number generator to choose which student should be the next to answer a question. Like in realistic classroom settings, several things co-occurred (Chan et al., 2021). The videos furthermore portrayed a wide variety of math-gender stereotype-reinforcing behaviors, which, especially in this combination, would have been difficult to record in a non-scripted setting (see also Gartmeier, 2014).

I found awareness of math-gender stereotype-reinforcing behaviors related to math-gender misconceptions with a small but significant effect in Study 2. This finding is similar to prior research, which found teachers' ability to detect and critically evaluate teacher behaviors and student-teacher interactions related to aspects of their professional teacher knowledge (Bergin et al., 2017; Kersting et al., 2012). Since both measures were correlated with a small but significant effect, assessing teachers' awareness of math-gender stereotype-reinforcing behaviors with classroom videos may complement a questionnaire tool like the MGMQ(-s). Such a close-to-behavioral measure can prevent impression management based on social desirability, which is common for self-report measures like questionnaires (van de Mortel, 2008). Such impression management is unlikely to occur for a tool that requires the

application of professional knowledge in subtle situations. Preventing social desirability is especially interesting for assessing the math-gender stereotype and its related constructs since it is socially undesirable (Agut et al., 2023).

Such a close-to-behavioral video measure furthermore holds the potential to bridge the theory-practice gap in teacher education by calling for the application of professional knowledge (Greenway et al., 2019; Grossmann, 2011; Kersting et al., 2012). Hence, apart from its diagnostic purpose, the videos could also be applied to teacher training to increase awareness of math-gender stereotype-reinforcing behaviors.

In this dissertation, I scripted and filmed four different classroom videos to assess awareness of a breadth of math-gender stereotype-reinforcing behaviors. Yet, the math-gender stereotype-reinforcing behaviors shown in the videos are not exhaustive of all potential math-gender stereotype-reinforcing behaviors in a classroom setting. Future studies could add to this body of research by applying different classroom videos with math-gender stereotype-reinforcing behaviors. Future videos could be matched to the items of the MGMQ(s-) to better assess the application of the specific, relevant knowledge or the detrimental influence of math-gender misconceptions on such knowledge. An example video to match the *empathizing-systemizing misconception* could be of a model teacher who praises a male student for their systematical thinking abilities while praising a female student for her empathetic abilities.

### ***Findings on Teachers' Math-Gender Misconceptions and Teachers' Awareness***

Since, to my best knowledge, there are neither measures that allow for the assessment of the three specific math-gender misconceptions of *empathizing-systemizing*, *girls' compensation*, and *girls' non-compensability*, nor measures that allow assessing awareness of the specific math-gender stereotype-reinforcing behaviors as portrayed in the above-described classroom videos, first respective findings regarding prevalence and awareness were in the focus of this dissertation. Across all three studies in this dissertation, I found math-gender misconceptions prevalent among pre-service and in-service teachers. About 35 % to almost 50 % of these (future) teachers held at least one math-gender misconception throughout all studies. Moreover, I found low average awareness, especially of implicit math-gender stereotype-reinforcing behaviors.

The findings called for interventions that both target teachers' math-gender misconceptions and, given the relationship of both constructs, potentially increase awareness of math-gender stereotype-reinforcing behaviors. Since *personalized refutation texts* have been researched to effectively reduce teachers' misconceptions (Dersch et al., 2022a), they appeared promising. I therefore researched them in the third study of this dissertation.

### ***Personalized Refutation Text***

Refutation texts, stating a commonly held misconception, a cue that the respective misconception is incorrect, and providing an alternative, scientifically-based explanation have been criticized for being overly authoritative (Zengilowski et al., 2021; 2022) and for not addressing learners' prerequisites (Asterhan & Dotan, 2018). In this dissertation, I addressed these aspects by applying *personalized* refutation texts, combining refutation text with personalized feedback, taking each learner's individual prior misconceptions into account (see also Dersch et al., 2022a). This dissertation aimed to specify if combining personalized feedback with refutation texts is necessary or if personalized feedback and expository text are sufficient for conceptual change. These questions were not sufficiently addressed in a prior study (Dersch et al., 2022a), which combined personalized feedback with an adaptive condition, providing a refutation text only if a misconception was prevalent. If teachers were uncertain or did not hold a misconception, they received an expository text. While the personalized, adaptive condition was most effective in said study, it remained unclear if this effect was due to personalization, potentially making the refutation obsolete.

Personalized feedback in itself can serve as a metacognitive prompt, similar to a refutation text (Vosniadou, 1994), by providing the information that a learner holds a misconception (Asterhan & Dotan, 2018). Furthermore such personalized refutation text can appear less prescriptive and authoritative (Zengilowski et al., 2021; 2022).

Hence, in Study 3 of this dissertation, I applied a 2x2 design comparing the effects of standard expository text, standard refutation text, personalized refutation text, and personalized expository text. I found that personalized expository text did not foster teachers' conceptual change as effectively as personalized refutation text. It is therefore beneficial to combine personalized feedback with refutation text. In line with theoretical assumptions (e.g., Strike & Posner, 1982), personalized refutation text seems to make the threshold between a learner's *own misconception* and alternative, correct conception more obvious because, different from personalized expository text, personalized refutation text additionally clearly states and disproves a misconception. However, in future studies, this must be researched using different samples and materials on different topics to draw robust conclusions.

### ***Personalized Refutation Texts Targeting Teachers' Math-Gender Misconceptions***

Since refutation text research tends to use the same materials in different experiments, this raises questions about the generalizability of the refutation text paradigm (Zengilowski et al., 2021; 2022). For teachers' conceptual change with *personalized* refutation texts, there is even less research, disallowing generalization. While a personalized (refutation) text intervention effectively refuted in-service teachers' multimedia learning misconceptions (Dersch et al., 2022a), it remained unclear if this effect was stable for other topics, such as math-gender misconceptions. Researching a potential generalization of the *personalized* refutation text effect with the topic of math-gender misconceptions was especially insightful for two reasons.

First, math-gender misconceptions are related to learners' identity, self-concept, and attitudes, for instance, to their math-gender stereotype, as found in Study 1 and Study 2 (see also Martin & Halverson, 1981), or to their academic self-concept (e.g., Arens et al., 2021; Sherman & Cohen, 2002). To protect the self, targeting and refuting misconceptions related to identity, self-concept, and attitudes can foster backfire effects. Backfire effects hinder the successful refutation of the targeted misconceptions or even reinforce them (Lewandowsky et al., 2020; Jacobson et al., 2021; Sherman & Cohen, 2002; Trevors et al., 2016).

Second, refuting teachers' math-gender misconceptions, especially in teachers as influential socializers (e.g., Gunderson et al., 2012), might help to improve gender equity in mathematics since math-gender misconceptions likely contribute to the manifestation of the math-gender stereotype (Study 1; Study2; see also Girelli, 2023; Rubel et al., 2024), which in turn contributes to female underrepresentation in mathematics (e.g., Cvencek et al., 2012; Greenwald et al., 2002; Eccles, 2011; Nosek & Smyth, 2011; Wang & Degol, 2017).

For participants who had held math-gender misconceptions in the pre-test, findings revealed that personalized refutation text about math-gender misconceptions was the most effective intervention. Findings suggest that personalized refutation text, among the text types researched here, is the most potent for changing teachers' math-gender misconceptions. The personalized refutation text effect seems to be consistent for different topics, such as multimedia learning (Dersch et al., 2022a) and gendered ability differences (Study 3). These findings take us a step closer to improved generalizability of the personalized refutation text effect and potentially contribute to narrowing the math-gender gap in the long run.

Additionally, to explore mechanisms lying behind the personalized refutation text effect, it was necessary to research affective and motivational processes related to conceptual change (see also Sinatra & Seyranian, 2015).

### *The Role of Negative Affect*

Previous research has identified negative affect of internal and external blame as related to conceptual change and attitude change (e.g., Dersch et al., 2022a; Nabi, 1999; Trevors et al., 2016). Such negative affect can be caused by detecting a threshold between one's own misconceptions and the scientifically supported conception (Dersch et al., 2022a; Gregoire, 2003; Strike & Posner, 1982). Negative affect of *internal* blame, such as guilt, occurs if the threshold is appraised as a *challenge* rather than a threat (Dersch et al., 2022a; Gregoire, 2003). A slight negative affect of internal blame can foster motivation for conceptual change and attitude change (Gregoire, 2003; Nabi, 1999; Strike & Posner, 1982). An increase in negative affect of internal blame accordingly mediated the effect of personalized (refutation) texts on conceptual change about multimedia learning in a prior study (Dersch et al., 2022a).

The relation of math-gender misconceptions to attitudes, self-concept, and identity (e.g., Martin & Halverson, 1981; Sherman & Cohen, 2002; Study 1; Study 2) may, however, impact the role of negative affect. Stress caused by detecting a threshold between one's misconceptions related to self-concept, identity, or attitudes and the scientifically supported conceptions is more likely appraised as a *threat* than a challenge (Gregoire, 2003; Sherman & Cohen, 2002; Trevors, 2016). Threat appraisal may cause negative affect of *external* blame instead of negative affect of internal blame, which hampers conceptual change instead of facilitating it (Trevors et al., 2016). In this dissertation, I researched if the role of negative affect of internal blame remained stable or if negative affect of external blame would play a role for the attempted refutation of math-gender misconceptions.

Unlike in prior research (Dersch et al., 2022a), negative affect of *internal* blame did not mediate the effects of personalized refutation texts on conceptual change. Personalized refutation texts did not increase negative affect of internal blame compared to the other conditions. Negative affect related to internal blame was, however, related to fostering conceptual change, as expected and similar to previous research (Dersch et al., 2022a).

Personalized refutation texts being the most effective condition refuting teachers' math-gender misconceptions, however, not mediated via negative affect of internal blame, might suggest that *cognitive* processes, rather than *affective* processes, improved conceptual change in the personalized refutation text condition. *Personalization* may have activated the learners' prior knowledge about themselves. Such prior knowledge activation about the rich schemata of the self improves the integration of the to-be-learned information in these prior knowledge structures, which benefits learning (Markus, 1977; Martin & Halverson, 1981). Furthermore, combining personalized feedback and refutation text may have enhanced cognitive co-activation of the misconceptions and the alternative explanation (Kendeo & O'Brien, 2014) and improved metacognition (Vosniadou, 1994). These processes are also researched to benefit conceptual change (Kendeo & O'Brien, 2014; Vosniadou, 1994; Danielson et al., 2024).

A similar amount of internal blame throughout all conditions may be due to the *topic* of math-gender misconceptions rather than the *text type* (personalized vs. standard, refutation vs. expository). Regardless of text type, math-gender misconceptions may have caused negative affect of internal or external blame among some participants depending on their attitudes, like their related math-gender stereotype (Martin & Halverson, 1981; Study 1; Study 2). Accordingly, I found no intervention to trigger stronger external blame than the other conditions. This finding is interesting since I had expected negative affect of external blame to mediate backfire effects. Even though negative affect, which can be related to backfire effects (e.g., Trevors et al., 2016) did not play a role here, open questions regarding backfire effects remain to be addressed.

### ***Knowledge Backfire Effects***

Since targeting a misconception and attempting to refute it can cause different kinds of backfire effects, which lead to hindered conceptual change or even an unintended reinforcement of the targeted misconception (Lewandowsky et al., 2020), I also researched potential backfire effects in the third study of this dissertation.

I found knowledge backfire effects, which were, however, unrelated to negative affect of external blame. They were likely *familiarity backfire effects*. While worldview backfire effects are related to, or mediated by, negative affect of external blame (Trevors et al., 2016), familiarity backfire effects are not (Lewandowsky et al., 2020). Familiarity backfire effects occur due to mentioning a misconception several times without the need for a refutation, which increases familiarity, leading to an unintended strengthening of the misconceptions (Lewandowsky et al., 2020). Among participants whose misconceptions were reinforced, backfire effects were strongest in the *personalized refutation text* condition. Participants in this condition were confronted with the misconceptions more frequently than in the other conditions. The pre-test and the post-test both comprised misconception statements, the personalized feedback and the refutation-cue also stated the misconception. This accumulated exposure may have increased familiarity. Further, most participants who backfired did not hold strong misconceptions in the pre-test. Hence, they experienced increased familiarity without the need for a refutation, which causes familiarity backfire effects (Lewandowsky et al., 2020).

These findings suggest that a refutation intervention should be tailored to the prevalence of the to-be-targeted misconceptions, for example, by applying an adaptive algorithm, which only provides a refutation text if a misconception is prevalent and, in other cases, provides an expository text. Expository text, other than the refutation text, does not mention the misconception before providing an alternative explanation, reducing the misconceptions' increase in familiarity and, like this, preventing or reducing familiarity backfire effects for participants who do not hold misconceptions. This adaptivity was applied in a prior study (Dersch et al., 2022a), which, however, disallowed conclusions about the combination of personalization and expository texts vs. refutation texts. In sum, findings from both studies (Dersch et al., 2022a; Study 3) now imply that while personalized expository text is insufficient to enhance teachers' conceptual change if misconceptions are prevalent, it might be a valuable alternative if no misconception is prevalent since it avoids or reduces familiarity backfire effects. While personalized refutation text is more effective, it should only be applied if misconceptions are prevalent to avoid such familiarity backfire effects.

I furthermore found backfire effects for personalized refutation text on awareness of math-gender stereotype-reinforcing behaviors. Despite being the most potent condition for reducing math-gender misconceptions and contrary to our expectations, personalized refutation texts led to less awareness than the other conditions. These findings are similar to those of Ferrero et al. (2020), who also found teachers' backfire effects in a close-to-behavioral measure for their most effective refutation condition. The *continued influence effect* may contribute to such backfire effects on behaviors or behavioral measures (Seifert, 2014). The continued influence effect describes that misconceptions influence behaviors even after their refutation (Seifert, 2014). Low source credibility contributes to the continued influence effect (Eckert et al., 2021). While source credibility causing condition effects is unlikely in this dissertation since the texts were highly standardized between conditions, future research should assess source credibility to account for it if necessary.

Another explanation might be an interaction of familiarity backfire effects and experimental design, more precisely, the order of measures. In the personalized refutation text condition, participants were confronted with the math-gender misconception most frequently. Some participants already experienced familiarity backfire effects in the first post-test (the MGMQ-s; see above). Assessing awareness of math-gender stereotype-reinforcing behaviors

after the post-test may have led to familiarity backfire effects, which were expressed in the awareness measure after the post-test comprised the misconception statements again. Future research should, therefore, control for the measures' positioning to avoid such interactions.

Despite order effects as a potential explanation, the backfire effects of personalized refutation texts on math-gender stereotype-reinforcing behaviors remain worrying. While it is promising that personalized refutation texts were most effective at refuting teachers' math-gender misconceptions, the related constructs of teachers' awareness of math-gender stereotype-reinforcing behaviors and teachers' math-gender stereotype, as well as other *attitudes* are just as relevant, since they all contribute to the manifestation of the math-gender gap.

### *Attitudes*

Stereotypes are the cognitive component of an attitude (Fiske, 1982), comprising generalized knowledge and beliefs about social groups (e.g., Hilton & van Hippiel, 1996). Accordingly, I found math-gender misconceptions related to the math-gender stereotype in the first two studies of this dissertation (see also Martin & Halverson, 1981).

I wondered whether targeting math-gender misconceptions would affect the related math-gender stereotype and specific attitudes about gender equity in mathematics. While attitude change via refutation texts is possible (Jacobson et al., 2021; Sinatra & Seyranian, 2015; Thacker et al., 2020), this does not occur in all cases of an attempted refutation. Sinatra and Seyranian (2015) postulate that for successful attitude change, misconceptions should not be very robust (Sinatra & Seyranian, 2015). Since, to my best knowledge, the influence of an attempted refutation of math-gender misconceptions on related attitudes has not been researched yet, I tried narrowing this research gap.

Regarding potential differences between personalized feedback vs. no personalized feedback and refutation text vs. expository texts, I found that, compared to no feedback, personalized feedback, among those participants who experienced knowledge backfire effects, led to more positive specific attitudes toward gender equity in math. Knowledge backfire effects might go hand in hand with the backfiring of the related negative attitudes (Jacobson et al., 2021; Sinatra & Seyranian, 2015). Personalized feedback, however, seems to have mitigated such attitude backfire effects, which indicates that it may function as a protective factor, in line with the assumption that personalization can reduce the perceived prescriptiveness and authority often criticized in refutation texts (Zengilowski et al., 2021; 2022). This finding highlights the potential of personalization.

Future research could expand the potential of personalization and adaptivity. Adaptive, personalized feedback, thus far, has mainly focused on learners' performance in a prior knowledge test rather than on their attitudes or affective states (Maier & Klotz, 2022). Especially in the context of refutation interventions on attitude-prone topics, it would be beneficial to tailor personalized feedback to attitudes, which might influence the reception of an intervention. This way, refutation interventions could not only be adapted to misconception prevalence (Dersch et al., 2022a) but also to negative attitudes related to the topic. A refutation instruction could be made 'softer' by containing concessive statements if the participant holds a negative attitude towards the refutational information. While such concessive statements did not enhance conceptual change in a prior study (Prinz et al., 2022), this may differ for more attitude-prone topics. An example for a concessive refutation text targeting math-gender misconceptions could read as follows:

"Many people believe that [math-gender misconception]. This assumption is, however, not in line with current evidence. *It is, however, understandable that such misconceptions are common since many people witness female underrepresentation in math-related fields and try to find an explanation for this underrepresentation.* However, it is not due to girls' or women's mathematical talent but due to societal

influences. [Explanation of social mechanism contributing to the math-gender gap as alternative explanation].”

Despite the promising findings and potentials regarding personalization, I found no condition effects of text type on the math-gender stereotype. In the third study, after an intervention, the math-gender stereotype no longer correlated with math-gender misconceptions, as in Study 1 and Study 2. The lack of connection between math-gender stereotype and misconceptions in Study 3 may be due to a successful reduction of the math-gender misconceptions, while the math-gender stereotype remained stable. More extensive interventions seem necessary to reduce the math-gender stereotype (e.g., Kollmayer et al., 2020; Priyashanta et al., 2021).

Since the math-gender stereotype draws on consolidated, often overgeneralized, knowledge based on socialization and prior experiences (Bodenhausen, 1990), future interventions could address such knowledge. Interventions could increase learners' association of *mathematics* with *female* by improving female representation and visibility in mathematics and math-intensive STEM fields (e.g., Dasgupta, 2011; Dasgupta et al., 2015). A related intervention I am currently researching is the combination of refutation texts with everyday life narratives, targeting experiential and socialization aspects of math-gender misconceptions and stereotypes.

### **Limitations and Further Research**

In this dissertation, I targeted and attempted to narrow different research gaps. In doing so, I constructed different measures (Study 1: MGMQ; Study 2: MGMG-s; video task for awareness of math-gender stereotype-reinforcing behaviors), investigated their relation to the math-gender stereotype and to one another (Study 1, Study 2, Study 3), and experimentally researched the effects and side effects of an intervention (Study 3). While I addressed open questions to specific research gaps in the previous section, there are also overarching, global limitations.

I conducted all three studies in this dissertation with substantial samples of pre-service and in-service teachers. Like other psychology research, I applied convenience sampling (Novielli et al., 2023). Thus, samples are likely affected by self-selection. I advertised Study 1 as related to mathematics and gender. While I did not advertise mathematics and gender as topics for Study 2 and Study 3, self-selection may have occurred via dropout. All three studies in this dissertation were online, which lowers the barrier to dropping out compared to laboratory settings (Hoerger, 2010). Once participants identified the topic, they may have dropped out if they disliked it. In Study 3, this dropout might, for instance, be responsible for participants' little variation in negative affect. Participants with strong negative affect because of the topic may have dropped out of the study. Affect-driven dropout may have obscured the potential role of negative affect of internal and external blame for conceptual change and backfire effects.

While samples in this dissertation cannot be considered representative, this lowered representativeness may have led to underestimating math-gender misconception prevalence, their relationships with other measures, and intervention effects rather than overestimating them. Participants who take part in studies dealing with the math-gender gap, due to the topic's relation with attitudes (e.g., Study 1; Study 2; Study 3; see also Fiske, 1982), are likely open to the topic or interested in it. Accordingly, in Study 1, participants indicated, on average, positive views about feminism and a regular engagement with feminism. Interested and open participants might hold *fewer* math-gender misconceptions than the population of teachers they ought to resemble. Fewer misconceptions among participants compared to the German teacher population (see also negative misconception scores, Study 1) could contribute to underestimating the relationship of misconceptions with other measures like awareness of math-gender stereotype-reinforcing behaviors (Field, 2024). The correlation was small in Study 2 and insignificant in Study 3. Moreover, in Study 3, personalized refutation texts were only more effective than other text types if both math-gender

misconceptions had been prevalent in the pre-test, which was the case for about a third of participants. The analyses with this subgroup revealing effects regardless highlights the reliability of results.

In order to improve generalizability, future research could use representative samples or conduct laboratory research about math-gender misconceptions without openly advertising the topic to avoid topic-based self-selection. This would, however, cause other self-selection biases because busy in-service teachers (Creagh et al., 2023) might be unlikely to participate in a study at a university laboratory. In this dissertation, conducting online studies with convenience sampling allowed for the assessment of substantial groups of in-service and pre-service teachers, yielding final samples between 289 and 336 participants.

A further limitation of the samples in this dissertation is that, while the research topic was *math*-gender misconceptions, not all participating in-service and pre-service teachers were (future) math teachers. Nevertheless, a substantial part of the three study samples were math teachers or pre-service teachers studying mathematics as one of their future school subjects. Additionally, I did not find differences regarding math-gender misconceptions between math teachers and teachers of other subjects in either of the studies ( $ps > .05$ ). Similarities in misconception prevalence may be due to teachers of all subjects being exposed to the same math-gender stereotypes and potential math-gender misconceptions based on shared socialization contexts, such as schools (Wentzel, 2014). In this dissertation, I therefore targeted all teachers as influential socializers (Gunderson et al., 2012; Wentzel, 2014), potentially conveying their math-gender stereotypes and math-gender misconceptions to their students (e.g., Bian, 2022; Beilock et al., 2010; Mischel, 1966; see also above). This holds true not only for math teachers but also for German teachers. If, for instance, a German teacher states: “Languages may not be for you, but as a boy, you may still be good at math,” this would reinforce not only the language-gender stereotype favoring girls (e.g., Retelsdorf et al., 2015) but also the math-gender stereotype favoring boys (Gunderson et al., 2012).

A last limitation regarding the samples is that in all three studies within this dissertation, said samples were predominantly female. The gender representation in the study samples does not resemble the gender distribution in German society. However, this imbalanced gender distribution, with most participants being female, resembles the gender distribution of teachers in Germany: The Federal Office for Statistics in Germany assessed teachers’ gender in the school year of 2019/2020 and found that 73.1 % of teachers in general education were female (Destatis, 2023). Furthermore, math-gender misconceptions and awareness of math-gender stereotype-reinforcing behaviors did not differ between the genders in either of the studies included in this dissertation ( $ps > .05$ ). That is likely the case because the math-gender stereotype and math-gender misconceptions are, as discussed before, conveyed to all members of our society.

A second important set of limitations concerns measurement and assessment. It is crucial to note that the self-constructed measures for math-gender misconceptions and awareness of math-gender stereotype-reinforcing behaviors are not comprehensive. They do not capture all possible math-gender misconceptions or every behavior that could reinforce the math-gender stereotype in a math classroom. Since research applying similar methods, to my best knowledge, is scarce, it is currently not possible to assess a broader spectrum of either math-gender misconceptions or awareness of different math-gender stereotype-reinforcing behaviors than those assessed in this dissertation. Future research could conduct an updated literature review to identify more math-gender misconceptions and expand the math-gender misconception questionnaire. Furthermore, research interested in awareness of math-gender stereotype-reinforcing behaviors could script and film more classroom videos portraying math-gender stereotype-reinforcing behaviors.

Another limitation regarding the measurement of math-gender misconceptions, awareness of math-gender stereotype-reinforcing behaviors, and the math-gender stereotype are potential order effects influencing their manifestation. Such order effects are likely when measuring socially desirable or easily primed constructs, like the math-gender stereotype, math-gender misconceptions, and awareness of math-gender stereotype-reinforcing behaviors (Agut et al., 2023; Hofmann et al., 2005; van de Mortel, 2008). In Study 2, participants watched a model teacher's *math-gender stereotype-reinforcing* behaviors before indicating their math-gender stereotype. Being confronted with such behaviors may have influenced how male they perceived mathematics when filling in the math-gender stereotype measure. Moreover, as discussed above, presenting math-gender misconceptions several times in Study 3 may have caused familiarity backfire effects for misconceptions and awareness (Lewandowsky et al., 2020).

For measures that can be easily primed, such as awareness of implicit math-gender stereotype-reinforcing behaviors and the explicit stereotype-measure, or measures that are prone to impression management based on social desirability, such as self-report measures about the math-gender stereotype or math-gender misconceptions (Agut et al., 2023; van de Mortel, 2008), measure order will always matter and may influence outcomes. Future research applying these measures should randomize the order of measures to control for such order effects. I have done so in a study not included in this dissertation. There, I also measured easily primed constructs influenced by impression management, like the explicit and the implicit math-gender stereotype (Dersch et al., submitted).

A further limitation concerns the intervention I conducted in this dissertation. It did not include a delayed post-test. There are mixed findings regarding refutation texts' stability over time (Zengilowski et al., 2021). Hence, it is important to apply a delayed post-test to determine how the effects of a personalized refutation text will develop. While some research on long-term conceptual change found backfire effects for refutation texts in a delayed post-test (Ferrero et al., 2020), other research finds stable advantages of refutation text over non-refutation text (e.g., Aguilar et al., 2020; Asterhan & Resnick, 2020). According to a qualitative study, refutation information can also be food for thought or a first nudge, fostering motivation to further engage with a topic, leading to evolving conceptual change (e.g., McLure et al., 2020). Theoretical assumptions about conceptual change also propose that conceptual change does not occur once at a time. Instead, refutational information can evolve in memory, for instance, by learners being continuously more successful at inhibiting a misconception while improving at activating the 'correct' alternative conception (Kendeou & O'Brien, 2014). Similar positive long-term effects might occur for attitudes. Hence, effects on the math-gender stereotype, if any, might only be detectable in the longer run. Therefore, future research should assess the long-term effects of personalized refutation texts on teachers' conceptual change and potential attitude change via delayed post-testing, especially for topics entrenched with attitudes, such as math-gender misconceptions with the math-gender stereotype.

Apart from the above-discussed limitations of this dissertation, an important topic has remained out of the scope of this research. Besides the math-gender misconceptions researched in this dissertation, other misconceptions are potentially associated with stereotypes, calling for research on whether associations between stereotypes and misconceptions are generalizable for other topics. One example is the math-race stereotype, associating mathematics with Whites and East Asians but not with Latinx and Black people (e.g., Starr, 2025). The *intersectionality* of different stereotypes, meaning people belonging to more than one minority group (e.g., Muslim and female) and therefore being stereotyped by different, sometimes overlapping, adverse stereotypes, should furthermore be considered in future research (Keller et al., 2023; Yuval-Davis, 2006).

### Conclusion and Implications for Practice

The problem of female underrepresentation in mathematics is still an important problem to be solved due to its negative impact on women who miss important career chances (Corbett & Hill, 2012; Good et al., 2012) as well as on society, which misses potentially qualified workers in mathematics and math-intensive STEM fields (Dasgupta, 2011). I aimed to contribute to ‘solving this gender equation’ by identifying and measuring math-gender misconceptions in teachers as important socializers. I repeatedly found math-gender misconceptions prevalent among pre-service and in-service teachers (Study 1; Study 2; Study 3). I also found them related to awareness of math-gender stereotype-reinforcing behaviors (Study 2), calling for interventions targeting math-gender misconceptions to potentially improve awareness. Therefore, I tested personalized refutation texts as an intervention. They delivered short-term effects in reducing teachers’ math-gender misconceptions. In order to determine their long-term effects, delayed post-tests are necessary. Furthermore, future interventions should expand the beneficial effects of digital learning settings by applying more adaptivity. To prevent familiarity backfire effects, they could, for instance, adapt the text type a learner receives to their misconceptions, only presenting a refutation text if necessary (see Dersch et al., 2022a). Adaptivity could also be applied to learners’ attitudes related to a topic, presenting ‘softer’ concessive refutation texts if learners hold strong negative attitudes about a topic. Digital interventions, apart from allowing for personalization and adaptivity, allow reaching many teachers as influential socializers with busy schedules (Creagh et al., 2023; Gunderson et al., 2012), which is a further benefit of such interventions.

While personalized refutation texts did not reveal a direct effect reducing the math-gender stereotype, the successful refutation of math-gender misconceptions might ultimately reduce the math-gender stereotype since debunking math-gender misconceptions subtracts an upholding factor from the ‘gender equation’ (see also Girelli, 2023; Rubel et al., 2024). Furthermore, personalized refutation texts targeting teachers’ math-gender misconceptions are not the only intervention to improve female representation in mathematics (see Priyashantha et al., 2021). Instead, refuting teachers’ math-gender misconceptions could and should be one puzzle piece in a set of interventions. Such a set of interventions may consist of informing affected students about the existence of the math-gender stereotype and its impact on education (Peretz-Lange et al., 2021; Vasilyeva et al., 2018), targeting and trying to reduce *students’* math-gender stereotype (Zhao et al., 2018), and of increasing subjective task value for mathematics in female learners (Gaspard et al., 2015). Interventions may also reduce the adverse effects of the math-gender stereotype via protective social factors, like female peers, female mentors, and female role models in mathematics and other math-intensive STEM fields (Dasgupta, 2011; Dasgupta et al., 2015; Dennehy & Dasgupta, 2017; Young et al., 2013). All these interventions potentially increase female representation in mathematics in the long run. Due to the global nature of the math-gender stereotype (e.g., Greenwald et al., 2002), these interventions should be combined to have a potential impact. Due to the math-gender stereotypes’ self-perpetuating nature (e.g., Eagly & Koenig, 2021; Haines et al., 2016), meaningful effects on its reduction might only be visible in future generations. Given the beneficial impact reducing the math-gender stereotype may hold (Corbett & Hill, 2012; Dasgupta et al., 2011; Good et al., 2012), these efforts are worth it.

### References

- Aguilar, S. J., Polikoff, M. S., & Sinatra, G. M. (2019). Refutation texts: A new approach to changing public misconceptions about education policy. *Educational Researcher*, 48(5), 263-272.  
<https://doi.org/10.3102/0013189X19849416>
- Arens, A. K., Jansen, M., Preckel, F., Schmidt, I., & Brunner, M. (2021). The structure of academic self-concept: A methodological review and empirical illustration of central models. *Review of educational research*, 91(1), 34-72. <https://doi.org/10.3102/0034654320972186>
- Ashmore, R. D., & Del Boca, F. K. (1979). Sex stereotypes and implicit personality theory: Toward a cognitive—Social psychological conceptualization. *Sex roles*, 5(2), 219-248. <https://doi.org/10.1007/BF00287932>
- Asterhan, C. S., & Dotan, A. (2018). Feedback that corrects and contrasts students' erroneous solutions with expert ones improves expository instruction for conceptual change. *Instructional Science*, 1-19.  
<https://doi.org/10.1007/s11251-017-9441-1>
- Asterhan, C. S., & Resnick, M. S. (2020). Refutation texts and argumentation for conceptual change: A winning or a redundant combination?. *Learning and Instruction*, 65, 101265.  
<https://doi.org/10.1016/j.learninstruc.2019.101265>
- Atanasova, S., Robin, N., & Brovelli, D. (2024). Interest, learning opportunities and teaching experience as predictors of professional vision in gender-sensitive physics education. *International Journal of Science Education*, 1-23.  
<https://doi.org/10.1080/09500693.2024.2406528>
- Avolio, B., Chávez, J., & Vílchez-Román, C. (2020). Factors that contribute to the underrepresentation of women in science careers worldwide: A literature review. *Social Psychology of Education*, 23(3), 773-794.  
<https://doi.org/10.1007/s11218-020-09558-y>
- Avitzour, E., Choen, A., Joel, D., & Lavy, V. (2020). *On the origins of gender-biased behavior: The role of explicit and implicit stereotypes*. National Bureau of Economic Research. <https://doi.org/10.3386/w27818>
- Baron-Cohen, S. (2004). *The essential difference*. Penguin UK.
- Bandura, A. (1977a). *Social Learning Theory*. Prentice Hall.
- Bandura, A. (1977b). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W H Freeman/Times Books/ Henry Holt & Co.
- Becker, J. C., & Swim, J. K. (2011). Seeing the unseen: Attention to daily encounters with sexism as way to reduce sexist beliefs. *Psychology of Women Quarterly*, 35(2), 227-242. <https://doi.org/10.1177/0361684310397509>
- Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences*, 107(5), 1860-1863.  
<https://doi.org/10.1073/pnas.0910967107>
- Bergin, C., Wind, S. A., Grajeda, S., & Tsai, C. L. (2017). Teacher evaluation: Are principals' classroom observations accurate at the conclusion of training?. *Studies in Educational Evaluation*, 55, 19-26.  
<https://doi.org/10.1016/j.stueduc.2017.05.002>
- Bian, L. (2022). Gender Stereotypes and Education. In: VanderLaan, D.P., Wong, W.I. (eds) *Gender and Sexuality Development. Focus on Sexuality Research*. Springer, Cham. [https://doi.org/10.1007/978-3-030-84273-4\\_9](https://doi.org/10.1007/978-3-030-84273-4_9)
- Bodenhausen, G. V. (1990). Stereotypes as Judgmental Heuristics: Evidence of Circadian Variations in Discrimination. *Psychological Science*, 1(5), 319-322. <https://doi.org/10.1111/j.1467-9280.1990.tb00226.x>

- Brookover, W. B., Thomas, S., & Paterson, A. (1964). Self-concept of ability and school achievement. *Sociology of education*, 271-278. <https://doi.org/10.2307/2111958>
- Canning, E. A., Muenks, K., Green, D. J., & Murphy, M. C. (2019). STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Science advances*, 5(2), eaau4734. <https://doi.org/10.1126/sciadv.aau4734>
- Canning, E. A., Ozier, E., Williams, H. E., AlRasheed, R., & Murphy, M. C. (2022). Professors who signal a fixed mindset about ability undermine women's performance in STEM. *Social Psychological and Personality Science*, 13(5), 927-937. <https://doi.org/10.1177/19485506211030398>
- Chaiken, S., & Ledgerwood, A. (2012). A theory of heuristic and systematic information processing. *Handbook of theories of social psychology*, 1, 246-266.
- Chan, K. K. H., Xu, L., Cooper, R., Berry, A., & van Driel, J. H. (2021). Teacher noticing in science education: do you see what I see?. *Studies in Science Education*, 57(1), 1-44. <https://doi.org/10.1080/03057267.2020.1755803>
- Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of educational research*, 63(1), 1-49. <https://doi.org/10.3102/00346543063001001>
- Chow, A., & Salmela-Aro, K. (2011). Task-values across subject domains: A gender comparison using a person-centered approach. *International Journal of Behavioral Development*, 35(3), 202-209. <https://doi.org/10.1177/0165025411398184>
- Cimpian, J. R., Lubienski, S. T., Timmer, J. D., Makowski, M. B., & Miller, E. K. (2016). Have gender gaps in math closed? Achievement, teacher perceptions, and learning behaviors across two ECLS-K cohorts. *AERA Open* 2:23328588416673617. <https://doi.org/10.1177/2332858416673617>
- Clinton, V. (2019). Reading from paper compared to screens: A systematic review and meta-analysis. *Journal of research in reading*, 42(2), 288-325. <https://doi.org/10.1111/1467-9817.12269>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203771587>
- Cook, J., & Lewandowsky, S. (2011). *The Debunking Handbook*. University of Queensland. <http://sks.to/debunk>
- Corbett, C., & Hill, C. (2012). Graduating to a Pay Gap: The Earnings of Women and Men One Year after College Graduation. *American Association of University Women*.
- Corpus, J. H., & Lepper, M. R. (2007). The effects of person versus performance praise on children's motivation: Gender and age as moderating factors. *Educational Psychology*, 27, 487-508. <https://doi.org/10.1080/01443410601159852>
- Crain, R. M. (1996). The influence of age, race, and gender on child and adolescent multidimensional self-concept. <https://doi.org/10.1080/10573560308222>
- Creagh, S., Thompson, G., Mockler, N., Stacey, M., & Hogan, A. (2023). Workload, work intensification and time poverty for teachers and school leaders: A systematic research synthesis. *Educational Review*, 1-20. <https://doi.org/10.1080/00131911.2023.2196607>
- 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>
- Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2012). Balanced identity theory. *Cognitive consistency: A fundamental principle in social cognition*, 157-177.

- Danielson, R. W., Jacobson, N. G., Patall, E. A., Sinatra, G. M., Adesope, O. O., Kennedy, A. A., ... & Sunday, O. J. (2024). The effectiveness of refutation text in confronting scientific misconceptions: A meta-analysis. *Educational Psychologist*, 1-25. <https://doi.org/10.1080/00461520.2024.2365628>
- Dasgupta, N. (2011). Ingroup experts and peers as social vaccines who inoculate the self-concept: The stereotype inoculation model. *Psychological inquiry*, 22(4), 231-246. <https://doi.org/10.1080/1047840X.2011.607313>
- Dasgupta, N., Scircle, M. M., & Hunsinger, M. (2015). Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. *Proceedings of the National Academy of Sciences*, 112(16), 4988-4993. <https://doi.org/10.1073/pnas.1422822112>
- Dennehy, T. C., & Dasgupta, N. (2017). Female peer mentors early in college increase women's positive academic experiences and retention in engineering. *Proceedings of the National Academy of Sciences*, 114(23), 5964-5969.
- Dersch, A.-S., Platte, S. J., Watt, H., & Eitel, A. (submitted). Thinking is not Telling: Student Teachers' Implicit Math-Gender Association is Unrelated to their Explicit Math-Gender Stereotype.
- Dersch, A.-S., Renkl, A., & Eitel, A. (2022a). Personalized refutation texts best stimulate teachers' conceptual change about multimedia learning. *Journal of Computer Assisted Learning*. <https://doi.org/10.1111/jcal.12671>
- \*Dersch, A.-S., Heyder, A., & Eitel, A. (2022b). Exploring the Nature of Teachers' Math-Gender Stereotypes: The Math-Gender Misconception Questionnaire. *Frontiers in Psychology*, 1728. <https://doi.org/10.3389/fpsyg.2022.820254> [Study 1]
- \*Dersch, A.-S., Heyder, A., & Eitel, A. (submitted). Teacher's Awareness of Math-gender Stereotype-reinforcing Behaviors: A Video Experiment. [Study 2]
- \*Dersch, A.-S., Thacker, I., & Eitel, A. (submitted). Should We Tell Teachers That They Are Incorrect? How Personalized Refutation Texts Promotes or Hampers Conceptual Change. [Study 3]
- Destatis (2023, July 4). Schulen: Lehrkräfte nach Schulart und Beschäftigungsumfang. *Datenreport 2021* [Schools: Teachers Regarding School Type and Scope of Employment. *Data report 2021*. [https://www.destatis.de/DE/Service/Statistik-Campus/Datenreport/Downloads/datenreport-2021-kap-3.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Service/Statistik-Campus/Datenreport/Downloads/datenreport-2021-kap-3.pdf?__blob=publicationFile)
- Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International journal of science education*, 25(6), 671-688. <https://doi.org/10.1080/09500690305016>
- Dweck, C. (1999). *Self-Theories: Their role in motivation, personality and development*. Taylor and Francis/Psychology Press.
- Dweck, C. (2015). Carol Dweck revisits the growth mindset. *Education Week*, 35(5), 20 – 24. <https://doi.org/10.1111/bjep.12072>
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt brace Jovanovich college publishers.
- Eagly, A. H., & Koenig, A. M. (2021). The vicious cycle linking stereotypes and social roles. *Current Directions in Psychological Science*, 30(4), 343-350. <https://doi.org/10.1177/09637214211013775>
- Eccles, J. S. (1994). Understanding women's educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *Psychology of women quarterly*, 18(4), 585-609. <https://doi.org/10.1111/j.1471-6402.1994.tb01049.x>
- Eccles, J. (2011). Gendered educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *International Journal of Behavioral Development*, 35(3), 195–201. <https://doi.org/10.1177/0165025411398185>

- Eccles, J. S., Barber, B., & Jozefowicz, D. (1999). Linking gender to educational, occupational, and recreational choices: Applying the Eccles et al. model of achievement-related choices. <https://psycnet.apa.org/doi/10.1037/10277-007>
- Eccles, J. S., & Wigfield, A. (2024). The development, testing, and refinement of Eccles, Wigfield, and colleagues' situated expectancy-value model of achievement performance and choice. *Educational Psychology Review*, 36(2), 51. <https://doi.org/10.1007/s10648-024-09888-9>
- Ecker, U. K., Swire, B., & Lewandowsky, S. (2014). Correcting misinformation—A challenge for education and cognitive science. <https://doi.org/10.7551/mitpress/9737.003.0005>
- Ecker, U. K., & Antonio, L. M. (2021). Can you believe it? An investigation into the impact of retraction source credibility on the continued influence effect. *Memory & Cognition*, 49, 631-644. <https://doi.org/10.3758/s13421-020-01129-y>
- Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of personality and social psychology*, 71(1), 5.
- Eitel, A., Prinz, A., Kollmer, J., Niessen, L., Russow, J., Ludäscher, M., Renkl, A., & Lindner, M. A. (2021). The Misconceptions About Multimedia Learning Questionnaire: An Empirical Evaluation Study With Teachers and Student Teachers. *Psychology Learning & Teaching*, 20(3), 420–444. <https://doi.org/10.1177/14757257211028723>
- Eliot, L. (2010). *Pink brain, blue brain: How small differences grow into troublesome gaps-and what we can do about it*. Simon and Schuster. <https://doi.org/10.14507/er.v0.1310>
- Ellemers, N. (2018). Gender Stereotypes. *Annual Review of Psychology*, 69, 275–298. <https://doi.org/10.1146/annurev-psych-122216-011719>
- Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, 136(1), 103-127. <https://doi.org/10.1037/a0018053>
- Endres, T., Vössing, C., Renninger, A. K., Eitel, A., & Renkl, A. (2025). Sustaining Focus When It's Hardest: Emotional Design Strengthens Sustained Learning, Especially in Contexts with Attractive Alternatives *Computers & Education*, <https://doi.org/10.1016/j.compedu.2025.105313>
- Escovar, E., Rosenberg-Lee, M., Uddin, L. Q., & Menon, V. (2016). The empathizing-systemizing theory, social abilities, and mathematical achievement in children. *Scientific reports*, 6(1), 1 – 11. <https://doi.org/10.1038/srep23011>
- Faust, M.W. 1992. Analysis of physiological reactivity in mathematics anxiety. PhD diss., Bowling Green State University, Ohio.
- Ferguson, L. E., & Bråten, I. (2022, October). Unpacking pre-service teachers' beliefs and reasoning about student ability, sources of teaching knowledge, and teacher-efficacy: A scenario-based approach. In *Frontiers in Education* (Vol. 7, p. 975105). Frontiers Media SA. <https://doi.org/10.3389/educ.2022.975105>
- Ferrero, M., Hardwicke, T. E., Konstantinidis, E., & Vadillo, M. A. (2020). The effectiveness of refutation texts to correct misconceptions among educators. *Journal of experimental psychology: applied*, 26(3), 411. <https://psycnet.apa.org/doi/10.1037/xap0000258>
- FitzGerald, C., Martin, A., Berner, D., & Hurst, S. (2019). Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: a systematic review. *BMC Psychol.* 7, 1–12. <https://doi.org/10.1186/s40359-019-0299-7>

- Fiske, S. T. (1982). Schema-triggered affect: Applications to social perception. In *Affect and cognition: 17th Annual Carnegie Mellon symposium on cognition* (pp. 55-78). Hillsdale: Lawrence Erlbaum.
- Field, A. (2024). *Discovering statistics using IBM SPSS statistics*. Sage publications limited.
- 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.
- Gartmeier, M. (2014). Fiktionale Videofälle in der Lehrerinnen-und Lehrerbildung [Fictional Videos for Teacher Education]. *Beiträge zur Lehrerinnen-und Lehrerbildung*, 32(2), 235-246. <https://doi.org/10.25656/01:13868>
- Gaspard, H., Dicke, A. L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2015). More value through greater differentiation: gender differences in value beliefs about math. *Journal of educational psychology*, 107(3), 663. <https://doi.org/10.1037/edu0000003>
- Girelli, L. (2023). What does gender has to do with math? Complex questions require complex answers. *Journal of Neuroscience Research*, 101(5), 679–688. <https://doi.org/10.1002/jnr.25056>
- Gjøvik, Ø., Kaspersen, E., & Farsani, D. (2023). Stereotypical images of male and female mathematics teachers. *Research in mathematics education*, 25(2), 178-193. <https://doi.org/10.1080/14794802.2022.2041471>
- Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of personality and social psychology*, 102(4), 700. <https://doi.org/10.1037/a0026659>
- Gottfredson, L. S. (1998). The general intelligence factor.
- Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological review*, 109(1), 3. <https://doi.org/10.1037/0033-295X.109.1.3>
- Gregoire, M. (2003). Is it a challenge or a threat? A dual-process model of teachers' cognition and appraisal processes during conceptual change. *Educational psychology review*, 15, 147-179. <https://doi.org/10.1023/A:1023477131081>
- Grossman, P. (2011). Framework for teaching practice: A brief history of an idea. *Teachers College Record*, 113(12), 2836-2843. <https://doi.org/10.1177/016146811111301205>
- Gunderson, E. A., Hamdan, N., Sorhagen, N. S., & D'Esterre, A. P. (2017). Who needs innate ability to succeed in math and literacy? Academic-domain-specific theories of intelligence about peers versus adults. *Developmental psychology*, 53(6), 1188. <https://psycnet.apa.org/doi/10.1037/dev0000282>
- Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The Role of Parents and Teachers in the Development of Gender-Related Math Attitudes. *Sex Roles*, 66(3-4), 153–166. <https://doi.org/10.1007/s11199-011-9996-2>
- Haines, E. L., Deaux, K., & Lofaro, N. (2016). The times they are a-changing... or are they not? A comparison of gender stereotypes, 1983–2014. *Psychology of Women Quarterly*, 40(3), 353-363. <https://doi.org/10.1177/0361684316634081>
- Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychol. Sci. Public Inter.* 8, 1–51. <https://10.1111/j.1529-1006.2007.00032.x>
- Hamilton, D., & Sherman, J. Stereotypes. In *Handbook of Social Cognition, Second edition* (Vol. 2, pp. 1–58). (Original work published 2014)
- Hand, S., Rice, L., & Greenlee, E. (2017). Exploring teachers' and students' gender role bias and students' confidence in STEM fields. *Social Psychology of Education*, 20(4), 929-945. <https://doi.org/10.1007/s11218-017-9408-8>

- Heyder, A., Steinmayr, R., and Kessels, U. (2019). Do teachers' beliefs about math aptitude and brilliance explain gender differences in children's math ability self-concept?. *Frontiers in Education*, 4(34), 1 – 11. <https://doi.org/10.3389/educ.2019.00034>
- Heyder, A., Weidinger, A. F., Cimpian, A., and Steinmayr, R. (2021). Teachers' belief that math requires innate ability predicts lower intrinsic motivation among low-achieving students. *Learning and Instruction*, 65, 101220. <https://doi.org/10.1080/10615809708249302>
- Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, 47, 237–271. <https://doi.org/10.1146/annurev.psych.47.1.237>
- Hinton, P. (2017). Implicit stereotypes and the predictive brain: cognition and culture in “biased” person perception. *Palgrave Communications*, 3(1). <https://doi.org/10.1057/palcomms.2017.86>
- Hoerger, M. (2010). Participant dropout as a function of survey length in Internet-mediated university studies: Implications for study design and voluntary participation in psychological research. *Cyberpsychology, Behavior, and Social Networking*, 13(6), 697 – 700. <https://doi.org/10.1089/cyber.2009.0445>
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality & Social Psychology Bulletin*, 31(10), 1369–1385. <https://doi.org/10.1177/0146167205275613>
- Holder, K., & Kessels, U. (2017). Gender and ethnic stereotypes in student teachers' judgments: A new look from a shifting standards perspective. *Social psychology of education*, 20, 471-490. <https://doi.org/10.1007/s11218-017-9384-z>
- Hu, L.T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Hughes, S., Lyddy, F., & Lambe, S. (2013). Misconceptions about Psychological Science: A Review. *Psychology Learning & Teaching*, 12, 20 - 31. <https://doi.org/10.2304%2Fplat.2013.12.1.20>
- Jacobson, N. G., Thacker, I., & Sinatra, G. M. (2021). Here's hoping it's not just text structure: The role of emotions in knowledge revision and the backfire effect. *Discourse Processes*, 59(1-2), 13-35. <https://doi.org/10.1080/0163853X.2021.1925059>
- Jamil, F. M., Larsen, R. A., & Hamre, B. K. (2018). Exploring longitudinal changes in teacher expectancy effects on children's mathematics achievement. *Journal for Research in Mathematics Education*, 49(1), 57-90. <https://doi.org/10.5951/jresmetheduc.49.1.0057>
- Jehle, A. M., Groeneveld, M. G., van de Rozenberg, T. M., & Mesman, J. (2024). The hidden lessons in textbooks: Gender representation and stereotypes in European mathematics and language books. *European Journal of Education*, 59(4), e12716. <https://doi.org/10.1111/ejed.12716>
- Jussim, L., Harber, K. D., Crawford, J. T., Cain, T. R., & Cohen, F. (2005). Social reality makes the social mind: Self-fulfilling prophecy, stereotypes, bias, and accuracy. *Interaction Studies*, 6(1), 85-102. <https://doi.org/10.1075/is.6.1.07jus>
- Kahn, S., & Ginther, D. (2017). *Women and STEM*. National Bureau of Economic Research. <https://doi.org/10.3386/w23525>
- Keller, C. (2001). Effect of Teachers' Stereotyping on Students' Stereotyping of Mathematics as a Male Domain. *J. Soc. Psychol.* 141, 165–173. <https://doi.org/10.1080/00224540109600544>

- Keller, L., Luedtke, O., Preckel, F., & Brunner, M. (2023). Educational Inequalities at the Intersection of Multiple Social Categories: An Introduction and Systematic Review of the Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy (MAIHDA) Approach. *Educational Psychology Review*, 35(1), 31. <https://doi.org/10.1007/s10648-023-09733-5>
- Kendeou, P., & O'Brien, E. J. (2014). The knowledge revision components (KReC) framework: Processes and mechanisms. <https://doi.org/10.7551/mitpress/9737.003.0022>
- Kerkhoven, A. H., Russo, P., Land-Zandstra, A. M., Saxena, A., & Rodenburg, F. J. (2016). Gender stereotypes in science education resources: A visual content analysis. *PLoS one*, 11(11), e0165037. <https://doi.org/10.1371/journal.pone.0165037>
- Kersey, A. J., Csumitta, K. D., & Cantlon, J. F. (2019). Gender similarities in the brain during mathematics development. *npj Science of Learning*, 4(1), 1—7. <https://doi.org/10.1038/s41539-019-0057-x>
- Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, 49(3), 568-589. <https://doi.org/10.3102/0002831212437853>
- Klineberg, O. (1951). The scientific study of national stereotypes. *International social science bulletin*, 3(3), 505-514.
- Kollmayer, M., Schultes, M. T., Lüftenegger, M., Finsterwald, M., Spiel, C., & Schober, B. (2020). REFLECT-a teacher training program to promote gender equality in schools. *Front. Educ.* 5:136. <https://doi.org/10.3389/educ.2020.00136>
- Lachance, J. A., & Mazzocco, M. M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learning and individual differences*, 16(3), 195—2016. <https://doi.org/10.1016/j.lindif.2005.12.001>
- Larkin, D. (2012). Misconceptions about “misconceptions”: Preservice secondary science teachers' views on the value and role of student ideas. *Science Education*, 96(5), 927-959. <https://doi.org/10.1002/sce.21022>
- Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E. Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., Vraga, E. K., Wood, T. J., Zaragoza, M. S. (2020). The Debunking Handbook 2020. <https://doi.org/10.17910/b7.1182>
- Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological science in the public interest*, 13(3), 106-131. <https://doi.org/10.1177/1529100612451018>
- Leslie, S. J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, 347(6219), 262-265. <https://doi.org/10.1126/science.1261375>
- Li, Q. (1999). Teachers' beliefs and gender differences in mathematics: a review. *Educational Research*, 41(1), 63–76. <https://doi.org/10.1080/0013188990410106>
- Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychological bulletin*, 136(6), 1123. <https://doi.org/10.1037/a0021276>
- Maheswaran, D., & Chaiken, S. (1991). Promoting systematic processing in low-motivation settings: Effect of incongruent information on processing and judgment. *Journal of personality and social psychology*, 61(1), 13. <https://doi.org/10.1037/0022-3514.61.1.13>
- Markus, H. (1977). Self-schemata and processing information about the self. *Journal of personality and social psychology*, 35(2), 63. <https://psycnet.apa.org/doi/10.1037/0022-3514.35.2.63>

- Martin, C. L., & Halverson, C. R. (1981). A schematic processing model of sex typing and stereotyping in children. *Child development, 1119-1134*. <https://doi.org/10.2307/1129498>
- Martingano, A. J., & Konrath, S. (2022). How cognitive and emotional empathy relate to rational thinking: empirical evidence and meta-analysis. *The Journal of Social Psychology, 162*(1), 143–160. <https://doi.org/10.1080/00224545.2021.1985415>
- Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2. Ed. pp. 43–71). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.005>
- McAfee, M., & Hoffman, B. (2021). The morass of misconceptions: How unjustified beliefs influence pedagogy and learning. *International Journal for the Scholarship of Teaching and Learning, 15*(1), 4. <https://doi.org/10.20429/ijstol.2021.150104>
- McFarland, L. A., Lev-Arey, D. M., & Ziegert, J. C. (2003). An examination of stereotype threat in a motivational context. *Human Performance, 16*(3), 181-205. [https://doi.org/10.1207/S15327043HUP1603\\_2](https://doi.org/10.1207/S15327043HUP1603_2)
- McLure, F., Won, M., & Treagust, D. F. (2020). ‘Even though it might take me a while, in the end, I understand it’: a longitudinal case study of interactions between a conceptual change strategy and student motivation, interest and confidence. *Disciplinary and Interdisciplinary Science Education Research, 2*, 1-17. <https://doi.org/10.1186/s43031-020-00027-2>
- Menz, C., Spinath, B., & Seifried, E. (2021). Where do pre-service teachers’ educational psychological misconceptions come from?. *Zeitschrift für Pädagogische Psychologie*. <https://doi.org/10.1024/1010-0652/a000299>
- Menz, C., Spinath, B., Hendriks, F., & Seifried, E. (2024). Reducing educational psychological misconceptions: How effective are standard lectures, refutation lectures, and instruction in information evaluation strategies?. *Scholarship of Teaching and Learning in Psychology, 10*(1), 56. [10.1037/stl0000269](https://doi.org/10.1037/stl0000269)
- Miller, D. I., Eagly, A. H., & Linn, M. C. (2015). Women’s representation in science predicts national gender-science stereotypes: Evidence from 66 nations. *Journal of Educational Psychology, 107*(3), 631–644. <https://doi.org/10.1037/edu0000005>
- Mischel, W. (1966). A social-learning view of sex differences in behavior. In E. E. Maccoby (Ed.), *The development of sex differences* (pp. 57 – 81). Stanford University Press.
- Moser, F., & Hannover, B. (2014). How gender fair are German schoolbooks in the twenty first century? An analysis of language and illustrations in schoolbooks for mathematics and German. *European journal of psychology of education, 29*(3), 387-407. <https://doi.org/10.1007/s10212-013-0204-3>
- Muenks, K., Canning, E. A., LaCosse, J., Green, D. J., Zirkel, S., Garcia, J. A., & Murphy, M. C. (2020). Does my professor think my ability can change? Students’ perceptions of their STEM professors’ mindset beliefs predict their psychological vulnerability, engagement, and performance in class. *Journal of Experimental Psychology: General, 149*(11), 2119–2144. <https://doi.org/10.1037/xge0000763>
- Muis, K. R., Sinatra, G. M., Pekrun, R., Winne, P. H., Trevors, G., Losenno, K. M., & Munzar, B. (2018). Main and moderator effects of refutation on task value, epistemic emotions, and learning strategies during conceptual change☆. *Contemporary Educational Psychology, 55*, 155-165. <https://doi.org/10.1016/j.cedpsych.2018.10.001>
- Munthe, E. (2001). Professional uncertainty/certainty: How (un) certain are teachers, what are they (un) certain about, and how is (un) certainty related to age, experience, gender, qualifications and school type?. *European journal of teacher education, 24*(3), 355-368. <https://doi.org/10.1080/02619760220128905>

- Nabi, R. L. (1999). A cognitive-functional model for the effects of discrete negative emotions on information processing, attitude change, and recall. *Communication theory*, 9(3), 292-320. <https://doi.org/10.1111/j.1468-2885.1999.tb00172.x>
- Nagy, G., Garrett, J., Trautwein, U., Cortina, K. S., Baumert, J., & Eccles, J. S. (2008). Gendered high school course selection as a precursor of gendered careers: The mediating role of self-concept and intrinsic value. <https://psycnet.apa.org/doi/10.1037/11706-004>
- Nandi, A., Halder, T., & Das, T. (2024). Gender Stereotypes and Bias in School Mathematics Textbooks: A Systematic Literature Review. *Current Perspectives in Educational Research*, 7(2), 36-50. <https://doi.org/10.46303/cuper.2024.8>
- Nelson, T. E., Acker, M., & Manis, M. (1996). Irrepressible stereotypes. *Journal of Experimental Social Psychology*, 32(1), 13-38. <https://doi.org/10.1006/jesp.1996.0002>
- Nosek, B. A., & Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *American Educational Research Journal*, 48(5), 1125-1156. <https://doi.org/10.3102%2F0002831211410683>
- Novielli, J., Kane, L., & Ashbaugh, A. R. (2023). Convenience sampling methods in psychology: A comparison between crowdsourced and student samples. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*.
- Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.).
- Nürnberger, M., Nerb, J., Schmitz, F., Keller, J., & Sütterlin, S. (2016). Implicit gender stereotypes and essentialist beliefs predict preservice teachers' tracking recommendations. *The Journal of Experimental Education*, 84(1), 152-174. <https://doi.org/10.1080/00220973.2015.1027807>
- Ommundsen, Y., Haugen, R., & Lund, T. (2005). Academic self-concept, implicit theories of ability, and self-regulation strategies. *Scandinavian Journal of Educational Research*, 49(5), 461-474. <https://doi.org/10.1080/00313830500267838>
- Pajares, F. (1997). Current directions in self-efficacy research. *Advances in motivation and achievement*, 10(149), 1-49.
- Passolunghi, M. C., Rueda Ferreira, T. I., & 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>
- Peretz-Lange, R., Perry, J., & Muentener, P. (2021). Developmental shifts toward structural explanations and interventions for social status disparities. *Cognitive Development*, 58, 101042. <https://doi.org/10.1016/j.cogdev.2021.101042>
- Prinz, A., Kollmer, J., Flick, L., Renkl, A., & Eitel, A. (2022). Refuting student teachers' misconceptions about multimedia learning. *Instructional Science*, 50(1), 89-110. <https://doi.org/10.1007/s11251-021-09568-z>
- Priyashantha, K. G., De Alwis, A. C., & Welmilla, I. (2023). Facets of Gender Stereotypes Change: A Systematic Literature Review. *Kelaniya Journal of Management*, 12(1). <https://doi.org/10.4038/kjm.v12i1.7689>
- Retelsdorf, J., Schwartz, K., & Asbrock, F. (2015). "Michael can't read!" – Teachers' gender stereotypes and boys' reading self-concept. *Journal of Educational Psychology*, 107, 186- 194. <https://doi.org/10.1037/a0037107>
- Riegle-Crumb, C., & Humphries, M. (2012). Exploring bias in math teachers' perceptions of students' ability by gender and race/ethnicity. *Gender & Society*, 26(2), 290-322. <https://doi.org/10.1177/0891243211434614>

- Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., & Copur-Gencturk, Y. (2014). Teachers' perceptions of students' mathematics proficiency may exacerbate early gender gaps in achievement. *Developmental psychology*, *50*(4), 1262. <https://doi.org/10.1037/a0035073>
- Rubel, L. H., Ayalon, M., & Shahbari, J. A. (2024). Sequencing & selecting solutions in a gendered world. *Mathematical Thinking and Learning*, *26*(2), 185-207. <https://doi.org/10.1080/1098606>
- Sáinz, M., Fàbregues, S., & Solé, J. (2020). Parent and teacher depictions of gender gaps in secondary student appraisals of their academic competences. *Frontiers in Psychology*, *11*, 573752. <https://doi.org/10.3389/fpsyg.2020.573752>
- Salmeron, L., Altamura, L., Delgado, P., Karagiorgi, A., & Vargas, C. (2024). Reading comprehension on handheld devices versus on paper: A narrative review and meta-analysis of the medium effect and its moderators. *Journal of Educational Psychology*, *116*(2), 153. <https://doi.org/10.1037/edu0000830>
- Sánchez, E., García-Rodicio, H., & Acuña, S. R. (2009). Are instructional explanations more effective in the context of an impasse? *Instructional Science*, *37*, 537–563. <https://doi.org/10.1007/s11251-008-9074-5>
- Seo, E., & Lee, Y. K. (2021). Stereotype threat in high school classrooms: How it links to teacher mindset climate, mathematics anxiety, and achievement. *Journal of youth and adolescence*, *50*(7), 1410-1423. <https://doi.org/10.1007/s10964-021-01435-x>
- Sherman, D. K., & Cohen, G. L. (2002). Accepting threatening information: Self-affirmation and the reduction of defensive biases. *Current Directions in Psychological Science*, *11*(4), 119–123. [10.1111%2F1467-8721.00182](https://doi.org/10.1111/2F1467-8721.00182)
- Sidi, Y., Ophir, Y., & Ackerman, R. (2016). Generalizing screen inferiority-does the medium, screen versus paper, affect performance even with brief tasks?. *Metacognition and Learning*, *11*, 15-33. <https://doi.org/10.1007/s11409-015-9150-6>
- Sinatra, G. M., & Seyranian, V. (2015). Warm change about hot topics: The role of motivation and emotion in attitude and conceptual change about controversial science topics. In *Handbook of educational psychology* (pp. 259-270). Routledge.
- Skinner, B. F. (1963). Operant behavior. *American Psychologist*, *18*(8), 503. <https://psycnet.apa.org/doi/10.1037/h0045185>
- Skinner, B. F. (1977). The experimental analysis of operant behavior. *Annals of the New York Academy of Sciences*, *291*(1), 374-385. <https://doi.org/10.1111/j.1749-6632.1977.tb53088.x>
- Song, J., Zuo, B., Wen, F., & Yan, L. (2017). Math-gender stereotypes and career intentions: An application of expectancy–value theory. *British Journal of Guidance & Counselling*, *45*(3), 328-340. <https://doi.org/10.1080/03069885.2017.1283680>
- 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>
- Starr, C. R. (2025). A latent profile analysis of gender and ethnic/racial stereotypes about who belongs in physical sciences, technology, engineering, and mathematics fields. *Journal of Diversity in Higher Education*. <https://psycnet.apa.org/doi/10.1037/dhe0000666>
- Starr, C. R., Gao, Y., Rubach, C., Lee, G., Safavian, N., Dicke, A. L., Eccles, J. S., & Simpkins, S. D. (2023). “Who’s Better at Math, Boys or Girls?”: Changes in Adolescents’ Math Gender Stereotypes and Their Motivational Beliefs from Early to Late Adolescence. *Education Sciences*, *13*(9), 866. <https://doi.org/10.3390/educsci13090866>

- Starr, C. R., & Simpkins, S. D. (2021). High school students' math and science gender stereotypes: relations with their STEM outcomes and socializers' stereotypes. *Soc. Psychol. Educ.* 24, 273–298. <https://doi.org/10.1007/s11218-021-09611-4>
- 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), 324-335. <https://doi.org/10.1007/s11199-010-9924-x>
- Stockard, J. (2006). Gender socialization. *Handbook of the Sociology of Gender*, 215-227. [https://doi.org/10.1007/0-387-36218-5\\_11](https://doi.org/10.1007/0-387-36218-5_11)
- Strack, F., & Deutsch, R. (2015). The duality of everyday life: Dual-process and dual system models in social psychology. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology, Vol. 1. Attitudes and social cognition* (pp. 891–927). American Psychological Association. <https://doi.org/10.1037/14341-028>
- Strike, K. A., & Posner, G. J. (1982). Conceptual change and science teaching. *European Journal of Science Education*, 4(3), 231-240. <https://doi.org/10.1080/0140528820040302>
- Sweller, J., Van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, 10, 251-296. <https://doi.org/10.1023/A:1022193728205>
- Thacker, I., Sinatra, G. M., Muis, K. R., Danielson, R. W., Pekrun, R., Winne, P. H., & Chevrier, M. (2020). Using persuasive refutation texts to prompt attitudinal and conceptual change. *Journal of Educational Psychology*, 112(6), 1085. <https://psycnet.apa.org/doi/10.1037/edu0000434>
- 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>
- Tiedemann, J. (2002). Teachers' Gender Stereotypes as Determinants of Teacher Perceptions in Elementary School Mathematics. *Educational Studies in Mathematics*, 50(1), 49–62. <https://doi.org/10.1023/A:1020518104346>
- Tippett, C. D. (2010). Refutation text in science education: a review of two decades of research. *Internat. J. Sci. Educ.* 8, 951–970. <https://10.1007/s10763-010-9203-x>
- Trevors, G. J., Muis, K. R., Pekrun, R., Sinatra, G. M., & Winne, P. H. (2016). Identity and epistemic emotions during knowledge revision: A potential account for the backfire effect. *Discourse Processes*, 53(5-6), 339-370. <https://doi.org/10.1080/0163853X.2015.1136507>
- Vasilyeva, N., Gopnik, A., & Lombrozo, T. (2018). The development of structural thinking about social categories. *Developmental Psychology*, 54(9), 1735. <https://dx.doi.org/10.1037/dev0000555>
- Vinacke, W. E. (1957). Stereotypes as Social Concepts. *The Journal of Social Psychology*, 46(2), 229–243. <https://doi.org/10.1080/00224545.1957.9714322>
- van de Mortel, T. F. (2008). Faking it: social desirability response bias in self-report research. *Australian Journal of Advanced Nursing*, 25(4), 40-48. <https://doi.org/10.3316/informit.210155003844269>
- von Aufschnaiter, C. & Rogge, C. (2015). Conceptual Change in Learning. In R. Gunstone (Ed.), *Encyclopedia of Science Education* (pp. 209-218). Springer.
- Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and instruction* 4, 45–69. [https://10.1016/0959-4752\(94\)90018-3](https://10.1016/0959-4752(94)90018-3)

- Vosniadou, S., & Mason, L. (2012). Conceptual change induced by instruction: A complex interplay of multiple factors. In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, & M. Zeidner (Eds.), *APA educational psychology handbook, Vol. 2. Individual differences and cultural and contextual factors* (pp. 221–246). American Psychological Association. <https://doi.org/10.1037/13274-009>
- Wang, M.-T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): current knowledge, implications for practice, policy, and future directions. *Educ. Psychol. Rev.* 29, 119–140. <https://doi.org/10.1007/s10648-015-9355-x>
- Wang, M. T., Degol, J., & Ye, F. (2015). Math achievement is important, but task values are critical, too: Examining the intellectual and motivational factors leading to gender disparities in STEM careers. *Frontiers in psychology*, 6, 36. <https://doi.org/10.3389/fpsyg.2015.00036>
- Watt, H. M. G. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th- through 11th-grade Australian students. *Child Development*, 75, 1556–1574. <http://doi.org/10.1111/j.1467-8624.2004.00757.x>
- Watt, H. M. G., Shapka, J. D., Morris, Z. A., Durik, A. M., Keating, D. P., & Eccles, J. S. (2012). Gendered motivational processes affecting high school mathematics participation, educational aspirations, and career plans: a comparison of samples from Australia, Canada, and the United States. *Developmental Psychology*, 48, 1594–1611. <https://doi.org/10.1037/a0027838>
- Wentzel, K. R. (2014). "Socialization in school settings," in *Handbook of socialization: Theory and research, Second Edition*, eds J. E. Grusec and P. D. Hastings (New York, NY: The Guilford Press), 251–275.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary educational psychology*, 25(1), 68-81. <https://doi.org/10.1006/ceps.1999.1015>
- Wilén, W. W. (2004). Refuting misconceptions about classroom discussion. *The Social Studies*, 95(1), 33-39. <https://doi.org/10.3200/TSSS.95.1.33-39>
- Young, D. M., Rudman, L. A., Buettner, H. M., & McLean, M. C. (2013). The influence of female role models on women's implicit science cognitions. *Psychology of women quarterly*, 37(3), 283-292. <https://doi.org/10.1177/0361684313482109>
- Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. University of California, Los Angeles.
- Yuval-Davis, N. (2006). Intersectionality and feminist politics. *European journal of women's studies*, 13(3), 193-209. <https://doi.org/10.1177/02F1350506806065752>
- Zhao, F., Zhang, Y., Alterman, V., Zhang, B., & Yu, G. (2018). Can math-gender stereotypes be reduced? A theory-based intervention program with adolescent girls. *Current Psychology*, 37, 612-624.
- Zengilowski, A., Schuetze, B. A., Nash, B. L., & Schallert, D. L. (2021). A critical review of the refutation text literature: Methodological confounds, theoretical problems, and possible solutions. *Educational Psychologist*, 56(3), 175-195. <https://doi.org/10.1080/00461520.2020.1861948>
- Zengilowski, A., Nash, B. L., Schuetze, B. A., & Schallert, D. L. (2022). Bringing Refutation Texts Back to Their Literacy Roots: What do Critical Literacy and Culturally Responsive Pedagogy Have to Teach us About Students' Conceptual Change?. *Literacy Research: Theory, Method, and Practice*, 71(1), 341-358. <https://doi.org/10.1177/23813377221109544>



# Exploring the Nature of Teachers' Math-Gender Stereotypes: The Math-Gender Misconception Questionnaire

Anna-Sophia Dersch<sup>1\*</sup>, Anke Heyder<sup>2\*</sup> and Alexander Eitel<sup>1</sup>

<sup>1</sup> Teaching and Learning With Media, Department of Educational Psychology, Institute for Psychology, Justus Liebig University Giessen, Giessen, Germany, <sup>2</sup> Department of Educational and Differential Psychology, Institute for Psychology, TU Dortmund University, Dortmund, Germany

## OPEN ACCESS

### Edited by:

Jan Retelsdorf,  
University of Hamburg, Germany

### Reviewed by:

Joan Barth,  
University of Alabama, United States  
David Reilly,  
Griffith University, Australia

### \*Correspondence:

Anna-Sophia Dersch  
anna-sophia.dersch@psychol.uni-  
giessen.de  
Anke Heyder  
anke.heyder@tu-dortmund.de

### Specialty section:

This article was submitted to  
Educational Psychology,  
a section of the journal  
Frontiers in Psychology

Received: 22 November 2021

Accepted: 21 March 2022

Published: 14 April 2022

### Citation:

Dersch A-S, Heyder A and Eitel A  
(2022) Exploring the Nature  
of Teachers' Math-Gender  
Stereotypes: The Math-Gender  
Misconception Questionnaire.  
Front. Psychol. 13:820254.  
doi: 10.3389/fpsyg.2022.820254

Stereotypes of girls having weaker mathematical abilities than boys (math-gender stereotypes) are one factor reducing women's representation in mathematics. Teachers, as powerful socializers, often hold math-gender stereotypes. Reducing math-gender stereotypes in (student) teachers thus may foster women's representation in mathematics. Yet knowing the stereotypes' underlying assumptions is crucial to reducing it. Do math-gender stereotypes reflect elaborate, disproven theories about gender differences in math, meaning *math-gender misconceptions*? And if so, which math-gender misconceptions are behind math-gender stereotypes? This is the focus of the present research. The relevant literature implies the existence of three distinct misconceptions: (1) *empathizing-systemizing* ("As girls think rather empathically and boys think rather systematically, boys are on average more talented in math than girls"), (2) *girls' compensation* ("To achieve equally good grades in mathematics, boys have to make less effort because they are more talented than girls are"), and (3) *girls' non-compensability* ("Despite their on average stronger effort, girls are normally less proficient in math than boys"). We assessed these misconceptions in a student teacher sample ( $N = 303$ ) using our newly developed *Math-Gender Misconceptions Questionnaire*. Our results offer support for the expected three-factor structure of math-gender misconceptions. All three math-gender misconceptions showed good to acceptable scale reliabilities. On average, preservice teachers did not hold (strong) math-gender misconceptions. But a subgroup of 48,2% of preservice teachers held at least one of the three misconceptions. The *empathizing-systemizing* misconception was the most prevalent (32.0%) among the three misconceptions. Descriptively, endorsing the math-gender stereotype correlated most strongly with the *empathizing-systemizing* ( $r = 0.43$ ) and the *girls' compensation misconception* ( $r = 0.44$ ). This may indicate that especially these two misconceptions partly underlie math-gender stereotypes. As a consequence, refutation instructions designed to reduce these misconceptions may be a promising method to weaken math-gender stereotypes. Further research is needed to investigate to what degree reducing the present misconceptions is related to reducing math-gender stereotypes. Hence, this study is the first one of a planned series of studies on the relation between math-gender misconceptions and math-gender stereotypes.

**Keywords:** stem education, misconceptions, questionnaire, teacher education, math-gender stereotypes

## INTRODUCTION

Stereotypes of girls having weaker mathematical abilities than boys (math-gender stereotypes) are widely prevalent in Western societies (Nosek et al., 2010). Math-gender stereotypes reduce girls' interest, motivation, and performance in math, and lead to women being less likely to pursue mathematical professions (e.g., Wang and Degol, 2017). Teachers, as powerful socializers, also endorse math-gender stereotypes (e.g., Gunderson et al., 2012). Reducing math-gender stereotypes in (student) teachers thus seems a promising way to foster the representation of girls and women in mathematics. However, to address these stereotypes effectively, we must know about their nature and underlying assumptions. Do math-gender stereotypes reflect elaborate, yet disproven, theories about gender differences in mathematical abilities, that is, *misconceptions* (e.g., Eitel et al., 2021)? And if so, which misconceptions about mathematical abilities exist? We aim to answer these questions in the present research. This is important, because such misconceptions do not dissipate on their own – instead, overcoming them requires specific instructions in teacher education (refutation texts; Eitel et al., 2019; Menz et al., 2021).

In the literature, we identified three potential misconceptions associated with math-gender stereotypes about mathematical abilities: First, boys are assumed to be inherently better in math, because they supposedly think more systematically than girls, whereas girls think more empathically (Baron-Cohen, 2005; for disprove, see Escovar et al., 2016). Secondly, girls are assumed to succeed as well as boys in math only because they are hardworking, whereas boys are simply talented. This belief was detected in teachers and other socializers (Tiedemann, 2002; Robinson-Cimpian et al., 2014; Sáinz et al., 2020). Thirdly, if mathematical abilities are perceived as fixed (for theories of *fixed* and *growth mindset* and their influence on learners, see Dweck, 1999; Gunderson et al., 2017) and girls are ascribed less mathematical talent, then girls would be unable to compensate for their poorer mathematical abilities. In this study, we developed the *Math-Gender Misconception Questionnaire (MGMQ)* to investigate to what degree these three potential misconceptions are empirically separable, present in a student teacher sample, and linked to related constructs such as fixed mindsets of math ability (e.g., Leslie et al., 2015). Note that this study is the first of a planned series of studies on the relation between math-gender misconceptions and math-gender stereotypes.

### Gender Stereotypes About Mathematical Abilities

There is evidence that girls' – and boys' mathematical abilities are not inherently different (Lachance and Mazzocco, 2006; Kersey et al., 2019). However, with age, math-gender differences favoring male students emerge in some countries (Else-Quest et al., 2010). These gender differences are relatively small compared to other performance differences (e.g., caused by economic status; Bloom et al., 2008). Further, such differences are usually found in older learners (e.g., Reilly et al., 2015)

already influenced by societal gender attitudes (Eliot, 2010). Accordingly, gender differences are mediated by sex-role identity and related to cultural opportunity structures for women (Reilly, 2012). Moreover, gender stereotypes about girls' and women's lesser abilities in science, technology, engineering, and math (STEM) are widely prevalent in Western countries (Nosek et al., 2010; Nosek and Smyth, 2011; Hand et al., 2017) and predict women's lower STEM engagement (Hyde et al., 1990; Halpern et al., 2007; see Nosek and Smyth, 2011; for similar findings on reading and boys, see e.g., Retelsdorf et al., 2015; Muntoni and Retelsdorf, 2018). In this vein, in many Western countries, women remain underrepresented in the mathematical professions (Wang and Degol, 2017). The societal stereotypes of girls' and women's lesser math abilities (*math-gender stereotypes*) influence children from an early age (e.g., Eliot, 2010). Math-gender stereotypes are conveyed by parents, peers, and teachers (see e.g., Hannover, 2008). As school is especially important for children's socialization (Wentzel, 2014), children are prone to being influenced by teachers' math-gender stereotypes. According to the *Model of Achievement Related Choices* (Eccles et al., 1983), teachers, as part of the cultural milieu, hold gender stereotypes including math-gender stereotypes (Eccles, 2011; Gunderson et al., 2012). Teachers have more positive attitudes about male students' math performance, overrate male students' mathematical abilities and have higher expectations regarding male students' mathematical success (Riegle-Crumb and Humphries, 2012; Robinson-Cimpian et al., 2014; for a literature review, see Li, 1999). Further, teachers attribute failure in math to a lack of talent among girls, but to a lack of effort among boys (Tiedemann, 2002), and demonstrate a gender bias when evaluating students' performance in an experimental setting (underrating equal performance outcomes if they assume female learners achieved them (Avitzour et al., 2020; see also Holder and Kessels, 2017). Furthermore, the Eccles et al. (1983) model proposes that teachers' beliefs and behaviors influence their students' own gender roles and stereotypes (see Eccles, 2011). Teachers' own math-gender stereotypes thus predict students' math-gender stereotypes (Keller, 2001). Although these math-gender stereotypes seem to have decreased in school children (e.g., Passolunghi et al., 2014), recent research still suggests that even primary school children hold the perception of math being male-typed (Miller et al., 2015). These stereotypes then influence students' expectation of success and subjective task value (e.g., in mathematics), which in turn influences students' achievement-related choices. Math-gender stereotypes of girls lead to girls tending to make academic choices against mathematics (see Eccles, 2011). Apart from academic choices, math-gender stereotypes influence girls' sense of identity. The idea of math being male-typed (Miller et al., 2015) leads to girls developing less interest or preference for math when forming their identity. Thus, girls do not engage further with math, as girls try to establish their identity as distinct from the boys' identity and from male-typed interests (Bian et al., 2017). All in all, math-gender stereotypes reduce girls' interest, motivation, and performance in math, and, ultimately, lead to women being less likely to pursue mathematical professions (e.g., Wang and Degol, 2017). Further, according to learning theories, girls (and

boys) learn to behave according to gender stereotypes because parents, *teachers* and peers reinforce them for doing so (Mischel, 1966; Hannover, 2008). This process of operant conditioning leads to girls' engaging less with math as teachers – due to their math-gender stereotype – reinforce girls less than they reinforce boys for engaging with math. Besides that, math-gender stereotypes influence girls' – and women's performance through social-psychological mechanisms such as *self-fulfilling prophecies* or *stereotype threat*. When societal stereotypes are activated, girls are more likely to behave in a way that fulfills societal stereotypes and expectations. For example, teachers implicitly expressing their math-gender stereotypes and thus treating girls differently may instigate a worse math performance [see *self-fulfilling prophecy* (Merton and Merton, 1968)]. Just the fear itself of negative judgment in light of the math-gender stereotype can cause a disruption leading to girls' performing worse in math [see *stereotype threat* (Steele and Aronson, 1995)]. This means that teachers, who – because of their math-gender stereotypes – expect girls to perform worse, in fact contribute to female learners actually performing worse in standardized math tests (Geis, 1993; Smith et al., 1999; Spencer et al., 1999).

Finally, as powerful socializers, teachers do not only endorse math-gender stereotypes, their math-gender stereotypes influence girls' math attitudes and performance negatively (Gunderson et al., 2012; Carlana, 2019). Reducing teachers' stereotypes may therefore represent a means to increase women's representation in mathematics. To weaken stereotypes, however, it is important to know about their nature and underlying assumptions, which is in the focus of this research.

### Interrelation of Math-Gender Stereotypes and Math-Gender Misconceptions

Math-gender stereotypes and misconceptions about math abilities based on gender (*math-gender misconceptions*) are two theoretically related but separable constructs.

Stereotypes are based on oversimplified, overgeneralized beliefs (Klineberg, 1951); for instance, beliefs that a certain group member has certain attributes because they are a member of a group (Greenwald et al., 2002). Thus, the math-gender stereotype is the over-simplified, overgeneralized belief of girls having weaker mathematical abilities because of their gender (Math-gender). Stereotypes are rarely fully refuted (FitzGerald et al., 2019; Kollmayer et al., 2020). This may be the case, because the specific reasoning or (mis-)conceptions behind a global stereotype are hard to grasp and therefore hard to target (e.g., by refutation texts; Tippett, 2010). Likewise, empirical evidence showing that math-gender stereotypes persist despite being incorrect (e.g., Gunderson et al., 2012) is paralleled by the scarcity of research on how instruction can overcome these math-gender stereotypes (Kollmayer et al., 2020). In this study, we want to explore the specific reasoning behind teachers' math-gender stereotypes to prospectively provide refutation instruction. More specifically, we want to know whether endorsing math-gender stereotypes is related to holding *math-gender misconceptions* – subjectively plausible, yet disproven, theories about gender differences in mathematical abilities (for

misconception definition, see Vosniadou, 1994; Chi and Roscoe, 2002; Hughes et al., 2013).

### Math-Gender Misconceptions

Previous research suggests the potential presence of *three* specific misconceptions underlying gender stereotypes about mathematical abilities.

The first potential misconception refers to the Empathizing-Systemizing theory (Baron-Cohen, 2005) to explain the assumption of boys' better inherent mathematical abilities compared to girls' inherent mathematical abilities. The prominent Empathizing-Systemizing theory assumes that biological determinants explain gender differences in math. The Empathizing-Systemizing theory states that, because pre-natal testosterone-exposure is higher in the male fetus than the female, boys develop more systematic thinking in relation to less empathic thinking. Because pre-natal testosterone-exposure is lower in girls than boys, girls develop less systematic thinking in relation to more empathic thinking. According to the Empathizing-Systemizing theory, girls' weaker systematic thinking leads to lower mathematical abilities (Baron-Cohen, 2005). This view, however, is very one-sided and excludes societal factors scientifically proven to be important (e.g., Hannover, 2008; Eliot, 2010; Eccles, 2011; Wang and Degol, 2017). Further, even though female participants in some research did exhibit a higher ratio of empathic to systematic thinking than did men and vice-versa (e.g., Greenberg et al., 2018), this ratio-difference did not predict mathematical performance, even when researched in a huge sample (Escovar et al., 2016). In addition, the idea of empathic thinking being negatively associated with systematic thinking is not very convincing, considering that both refer to the construct of general thinking abilities [general intelligence (g); Gottfredson, 1998]. Consequently, the Empathizing-Systemizing theory itself represents a math-gender misconception (*empathizing-systemizing* misconception).

The second potential misconception, termed *girls' compensation*, refers to the belief that girls achieve similar math results as boys because they are hardworking, whereas boys are simply talented. However, girls actually report *less* intrinsic motivation in math than boys (e.g., Skaalvik and Rankin, 1994; Rodriguez et al., 2020; Heyder et al., 2020). As motivation is a strong predictor for effort and persistence (Skaalvik et al., 2015), girls are likely to be less driven to succeed in math. Girls are therefore very unlikely to achieve similar math results as boys only because they work harder. Furthermore, results from various studies suggest a similar level of mathematical talent in boys and girls: At a young age, girls and boys reveal gender similarities – rather than differences – in neural functioning when engaging with mathematical content (Kersey et al., 2019). In a longitudinal observation of primary school children (Lachance and Mazzocco, 2006), sex differences in math performance measured via standardized tests were minimal to non-existent. These empirical results offer no support for the idea that girls have lower math abilities overall. *Girls' compensation* thus counts as a math-gender misconception.

The third potential misconception, termed *girls' non-compensability*, also refers to the belief about gender differences

**TABLE 1** | Descriptive values for misconception items and scale reliabilities of the MGMQ.

<b>Empathizing-systemizing (ES):</b> $\omega = 0.88$ ; asymptotic $\omega = 0.90$	<b>Agreement rates</b> (min. = 0, max. = 1)	<b>Response certainty</b> (min. = 0, max. = 4)	<b>Misconception score<sup>a</sup></b> (min. = -4, max. = +4)	<b>Item-total correlation</b> (min. = 0, max. = 1)
ES1: <i>As girls think rather empathically and boys think rather systematically, boys are on average more talented in math than girls</i>	0.32	2.50 (0.94)	-1.30 (2.33)	0.57
ES2: <i>Mathematical relationships are usually easier to understand for boys than girls, because boys think in more systematic contexts</i>	0.39	2.31 (0.92)	-0.65 (2.40)	0.74
ES3: <i>As boy, more likely think in systematic categories, they fulfill more cognitive prerequisites for math than girls do</i>	0.39	2.22 (1.00)	-0.73 (2.33)	0.75
ES4: <i>Female empathy makes it easier for girls to deal with people, while boys are usually more gifted in systematic thinking and thus in math</i>	0.49	2.50 (0.97)	-0.28 (2.67)	0.77
ES5: <i>On average, girls think more empathically than boys do, while boys are more talented in systematic thinking and thus also in math</i>	0.44	2.34 (0.99)	-0.53 (2.49)	0.81
<b>Girls' compensation (GC):</b> $\omega = 0.76$ ; asymptotic $\omega = 0.91$				
GC1: <i>Mathematical content often comes easily to boys, while girls on average have to make more effort</i>	0.14	2.76 (0.86)	-2.16 (1.93)	0.58
GC2: <i>Girls normally have to work harder to perform as well in math as boys</i>	0.23	2.63 (0.83)	-1.60 (2.25)	0.61
GC3: <i>Girls compensate for their usually less aptitude in math compared to boys by being more diligent</i>	0.48	2.36 (0.91)	-0.19 (2.52)	0.46
GC4: <i>Girls usually need additional help to perform on par with boys in math</i>	0.14	2.61 (0.98)	-1.96 (1.98)	0.54
GC5: <i>To achieve equally good grades in math, boys have to make less effort because they are more talented than girls are</i>	0.17	2.67 (0.98)	-1.97 (2.05)	0.71
<b>Girls' non-compensability (GN):</b> $\omega = 0.72$ ; asymptotic $\omega = 0.68$				
GN1: <i>Since girls are on average less mathematically gifted, they should be assessed with different criteria than boys</i>	0.05	3.34 (0.87)	-3.10 (1.53)	0.56
GN2: <i>Girls should be rewarded with good grades for their stronger efforts in math, as they are not naturally as good at math as boys</i>	0.08	3.08 (0.98)	-2.74 (1.71)	0.62
GN3: <i>If the top of the class in math is a boy, it is because, in addition to his effort, he possesses a natural talent in math that diligent girls often lack</i>	0.18	2.80 (1.01)	-2.08 (2.14)	0.47
GN4: <i>Girls cannot fully compensate for their lack of aptitude for math with their on average greater diligence</i>	0.14	2.72 (0.89)	-2.11 (1.93)	0.45
GN5: <i>Despite their on average stronger effort, girls are normally less proficient in math than boys</i>	0.21	2.56 (0.97)	-1.67 (2.17)	0.43
<b>All items:</b> $\omega = 0.82$ ; asymptotic $\omega = 0.69$				

Agreement rates represent the proportion of participants agreeing statement. Descriptive values for response certainty and misconception scores represent means and standard deviations (in parentheses).

<sup>a</sup>Calculated by converting agreement into +1 and disagreement into -1, then multiplied with response certainty.

in mathematical talent. However, here the focus is on innate differences in mathematical talent that girls cannot compensate for later in life, because talent is assumed to be fixed. This *fixed mindset* is especially common in mathematics and other STEM subjects (e.g., Leslie et al., 2015; Gunderson et al., 2017; Canning et al., 2019) and also identified among teachers (Heyder et al., 2020). A fixed mindset stands in opposition to evidence of educational achievement, such as the *growth mindset* proposed by Dweck (1999, 2015). Accordingly, rather than being fixed, skills can improve over time with practice. However, people who hold the *girls' non-compensability* misconception assume that talent is fixed, and simultaneously ascribe girls less mathematical talent. In so doing, they assume girls cannot compensate for inherent talent differences in mathematical abilities. However, as described before, there is no evidence supporting the idea of girls having lower innate math abilities. Furthermore, the combination of a fixed mindset

and lack-of-talent assumptions is especially detrimental for female students' math-attitudes (Dweck, 2015; Heyder et al., 2019, 2020; Muenks et al., 2020) and for their performance (Canning et al., 2021).

## Current Study and Hypotheses

In this study, we present the newly developed *Math-Gender Misconception Questionnaire (MGMQ)* to assess teachers' misconceptions about gender differences in mathematics abilities. These misconceptions may underlie stereotypical thinking and behavior (see section "Interrelation of Math-Gender Stereotypes and Math-Gender Misconceptions"). By means of this questionnaire, we investigated to what degree the three potential misconceptions (*empathizing-systemizing*, *girls' compensation*, *girls' non-compensability*) are (1) empirically separable (*structure hypothesis*) and measurable by reliable scales, (2) present in a student teacher sample (*prevalence*

*hypothesis*), and (3) linked to theoretically related constructs (*association hypothesis*).

### Structure Hypothesis

We expect the MGMQ to assess three empirically separable, yet positively interrelated misconceptions. All three of the previously described misconceptions (see section “Interrelation of Math-Gender Stereotypes and Math-Gender Misconceptions”) are related to beliefs about gender differences in mathematical talent. Nevertheless, each misconception focuses on a different aspect: The *empathizing-systemizing* misconception provides an over-simplified explanation for the existence of gender differences in mathematical talent. The *girls’ compensation* misconception refers to girls managing to compensate for their lesser mathematical talent by investing effort. The misconception of *girls’ non-compensability* puts girls’ un-ability to compensate for their lack of talent into focus. Therefore, we expected the MGMQ data to fit a three-factor structure of math-gender misconceptions better than a general-factor structure with one homogeneous misconception construct in a confirmatory factor analysis.

### Prevalence Hypothesis

We expect student teachers to rather endorse the first two of the three potential misconceptions. Given the high prominence and face validity of the idea that girls think more empathically whereas boys think more systematically (Empathizing-Systemizing theory; Baron-Cohen, 2005), some student teachers may also believe that these thinking differences are related to worse mathematical abilities – a misconception (*empathizing-systemizing* misconception). Further, we expect some student teachers to endorse the *girls’ compensation* misconception referring to the belief that girls only succeed in math because they work hard, whereas boys who succeed are talented. This belief is likely to exist among teachers, because teachers attribute girls’ better math grades than boys’ math grades to the girls’ greater effort (Sáinz et al., 2020). Further, teachers perceive girls only as similarly math-competent as boys if girls work harder (Robinson-Cimpian et al., 2014). Likewise, teachers attribute girls’ weak mathematical performance to lacking talent, and boys’ weak mathematical performance to lacking effort (Tiedemann, 2002). This research also suggests that (student) teachers may endorse the *girls’ non-compensability* misconception to a lesser degree than the *girls’ compensation* misconception.

### Association Hypothesis

We first expect the three math-gender misconceptions to relate positively with the common math-gender stereotype found in previous research using a simple *female-to-male*-rating for math (for a similar measure, see Nosek, 2007; Nosek et al., 2010). We expect this association, as there are similarities and overlaps amongst math-gender stereotypes and math-gender misconceptions (Klineberg, 1951; Chi and Roscoe, 2002; Kollmayer et al., 2020). More specifically, we expect math-gender stereotypes to be partly based on math-gender misconceptions, which should be expressed in a moderate to high correlation

between the two. Secondly, we expect that holding the girls’ non-compensability misconception will relate positively with holding fixed-ability mindsets for mathematics (Leslie et al., 2015). Holding the girls’ non-compensability misconception means assuming that girls’ lack of talent cannot be compensated for, and is thus fixed. This misconception is similar to the idea of fixed ability mindsets for mathematics.

## MATERIALS AND METHODS

### Participants and Recruiting

A total of 303 student teachers [242 women, 61 men,  $M_{\text{age}} = 21.73$  ( $SD = 4.7$ , range = 18–51 years)] completed our online survey without dropping out. These data sets were complete (no missing data amongst them). The student teachers had studied on average for 2.28 semesters ( $SD = 2.28$ , range = 2–16 semesters). The student teachers’ school subjects were mostly German ( $n = 146$ ) and math ( $n = 118$ ), followed by other common subjects (e.g., English, biology, politics and economics, philosophy, geography, languages such as French, Spanish, or Latin). More than half of the participants (168; 55.5%) studied at least one STEM subject. Participants were studying to teach at the elementary ( $n = 79$ ) or secondary school level ( $n = 191$ ). Some participants were studying to teach in vocational education ( $n = 7$ ) or special needs education ( $n = 63$ ). Participants from all over Germany took part in this study; most were from Hessen. The participants, on average, held positive views about gender equality and feminism ( $M = 3.61$ ,  $SD = 0.84$ ; scale of 1 = not at all to 5 = very).

The communicated topic of the study was “Mathematics and Gender.” The online survey completion was possible between May and July of 2021. We recruited participants via teacher education lectures and seminars as well as via acquaintances. In total, 360 people clicked on the survey link, of which 303 participants (84.2%) completed the survey. Two people declined consent; the other 55 participants (15.3%) dropped out during the study and were not included in our analyses, yielding the final sample of 303 student teachers.

### Study Instruments

#### Math-Gender Misconception Questionnaire

The self-developed Math-Gender Misconceptions Questionnaire (MGMQ; see Table 1 for an English translation of the misconception items and Supplementary Appendix A for the German original containing all items) served as our main study instrument. It consisted of 30 items. These items comprised statements that participants first must answer with “I disagree” or “I agree” (i.e., verification). Second, each statement comprised a five-point Likert-scale assessing the participants’ certainty of having correctly responded to the current statement. The answer options were *very certain*, *certain*, *somewhat certain*, *uncertain*, *very uncertain* (i.e., certainty rating). Certainty ratings were horizontally aligned and presented below the corresponding verification part (see Figure 1, for an example item). These two ratings per item are crucial for assessing misconceptions: Holding a misconception should reflect in *incorrect* answers made with a (relatively)

As girls think rather empathically and boys think rather systematically, boys are on average more talented in mathematics than girls.

I disagree  I agree

How certain are you regarding your assessment of this statement?

very uncertain  uncertain  somewhat certain  certain  very certain

**FIGURE 1** | Example item.

*high certainty*. Incorrect answers with low certainty would rather reflect missing conceptions (see Eitel et al., 2021, for the argumentation). Of the 30 items in the MGMQ, 15 items targeted math-gender misconceptions (see Table 1) and 15 items were filler items. Of the 15 misconception items, always five referred to each of the three hypothesized misconceptions (*empathizing-systemizing*, *girls' compensation*, *girls' non-compensability*). The correct answer was to disagree with the misconception items.

The misconception items asked for all of the characterizing aspects of each hypothesized math-gender misconception, by also referring to research findings (Dweck, 1999; Tiedemann, 2002; Muenks et al., 2020; Sáinz et al., 2020), and academic as well as non-academic resources (Baron-Cohen, 2005; Escovar et al., 2016). For the *empathizing-systemizing* misconception, we constructed each of the items to address the combination of the following two stereotypical beliefs: (1) boys are better at math than girls (2) because boys think more systematically, whereas girls think more empathically. We constructed such complex items because only the combination of the two stereotypical beliefs [(1) gender differences in empathizing-systemizing and (2) their direct relation to mathematics performance] is a misconception. An example item was “As girls think rather empathically and boys think rather systematically, boys are on average more talented in mathematics than girls.” The same rationale for constructing items applies to the two other misconceptions.

For the *girls' compensation* misconception, items focused both on (1) the belief about gender differences in math talent and on (2) the beliefs that either girls compensate for their fewer talent through hard work or teachers compensate for girls' fewer talent by treating them differently than boys (e.g., more support). An example item regarding *girls' compensation* was “To achieve equally good grades in math, boys have to make less effort because they are more talented than girls are.”

For the *girls' non-compensability* misconception, items focused on both (1) the belief about girls being unable to compensate for their lack of talent even with hard work and (2) the belief about implications of this non-compensability in the treatment of genders (such as grading the girls more generously). An example item regarding *girls' non-compensability* was “Despite their on average stronger effort, girls are normally less proficient in math than boys.”

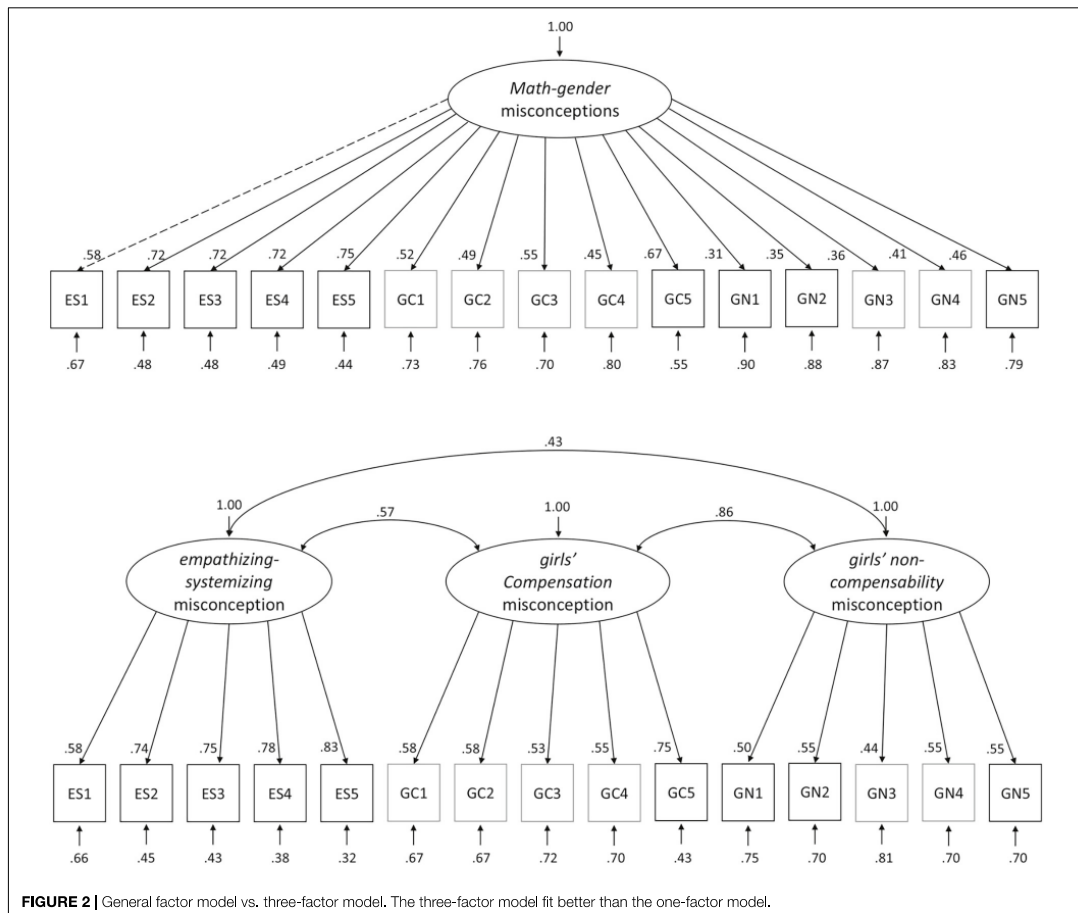
We intentionally formulated the misconception items as false statements to gain direct information as to whether the student teachers endorsed this particular misconception. Specifically, disagreeing with a correct statement (“the Earth is a sphere”) does not give direct information regarding the underlying misconception (the Earth could be flat, rectangular, a semi-sphere, etc.), whereas agreeing with the incorrect statement (“the Earth is flat”) provides direct information about endorsing this particular misconception (cf. Eitel et al., 2021).

The remaining 15 filler-items described true statements related to the math-gender gap, thus they were not misconceptions. An example filler item was “Amongst girls, math is more disliked than amongst boys.” The correct answer was to agree with these filler items. The filler-items served to balance the questionnaire. In total, 50% of the statements in the questionnaire were true (i.e., filler-items), while the other half of statements was untrue (i.e., misconception-items). We balanced the questionnaire in order to minimize response biases in the form of acquiescence tendencies (Moosbrugger and Kelava, 2012) because participants might think “some statements must be true” and answer accordingly (cf. Eitel et al., 2021).

Prior to inclusion in the questionnaire, an expert on the math-gender gap and an expert on developing questionnaires revised all items. Additionally, a four-member expert panel (one professor, two postdoctoral researchers, and a Ph.D. student from educational psychology) discussed and refined the questionnaire. Furthermore, we evaluated a prior version of this questionnaire within a pilot study with 246 student teachers. Results of this pilot study suggested that not one unitary construct of math-gender misconceptions but three misconceptions scales might best explain the questionnaire responses, namely the scales of *empathizing-systemizing*, *girls' compensation*, and *girls' non-compensability*. Based on these preliminary findings, we constructed the MGMQ with 15 misconception items, as the former version did not have sufficient misconception items per scale.

### Other Instruments

Furthermore, we assessed math-gender stereotypes similar to previous research (Nosek et al., 2010) by asking participants to indicate whether they perceived math as female or male. We used only part of the measure applied by Nosek et al. (2010), who assessed implicit and explicit math-gender stereotypes



**FIGURE 2 |** General factor model vs. three-factor model. The three-factor model fit better than the one-factor model.

together with liberal arts-gender stereotypes. Additionally, we extended the scale range to 9 answer options, starting from 1 (“very female”) via 5 (“neutral”) to 9 (“very male”), to potentially increase variance. The results nevertheless revealed that answers of 1 (“very female”), 8 and 9 (“very male”) were outliers in the answer distribution. We thus winsorized the distribution to reduce the biasing effect of the outliers in the correlational analyses.

We then assessed participants’ feminism using three items with five-tier Likert-scales each (from *not at all* to *very*). An example item was: “How important is the equality of the genders to you?” The internal consistency of the scale was good ( $\omega = 0.79$ ). For all three items in German and English, see **Supplementary Appendix B**.

We also assessed teachers’ *fixed mindset about math ability* with two items adapted from Leslie et al. (2015) and Heyder et al. (2020). An example item was: “Being among the best in math requires a special aptitude that just cannot be taught.” Both items

were highly correlated ( $r = 0.66, p < 0.001$ ) so that we calculated the mean score of both items ( $M = 4.07, SD = 1.42$ ).

Before ending the study, participants filled in their demographics such as age, sex, gender, mother language, study subjects, school type they will teach at or already teach at, and semesters studied.

## Procedure

When clicking on the web link, participants initially read about the voluntary nature of their participation, that they could end the study whenever they wanted without facing disadvantages, and that we would store all data for 10 years anonymously for the purpose of research only. Participants then gave their informed consent. Participants then read the instruction for the misconception questionnaire, which they then filled in. Then, participants rated how they perceived mathematics on a 9-tier Likert-scale (*female* to *male*). Afterward, participants filled in two items each on fixed mindset in math. They also indicated

their attitude toward feminism. Participants then provided basic demographic information. After participation, we thanked the participants and provided a full debriefing text. Participants took on average 13:44 min ( $SD = 5:48$  min) to complete the survey.

### Scoring the Misconceptions

We calculated misconception scores by multiplying agreement (coded with +1)/disagreement (coded with -1) and response certainty (coded from 0 = *very uncertain* to 4 = *very certain*; see Eitel et al., 2021). Thereby, we accounted for the nature of misconceptions: Misconceptions are incorrect and are subjectively highly plausible. Thus, if the person assumes an incorrect statement to be more plausible, this person endorses that statement more strongly, reflecting in higher certainty (see Eitel et al., 2021). This stronger endorsement of a misconception is reflected in higher misconception scores (see Table 1, for descriptive values). Participants who were very uncertain about an answer (coded with 0), regardless of whether it was correct or not ( $\pm 1$ ), got a misconception score of 0 (i.e.,  $\pm 1 \times 0 = 0$ ), because their (dis-)agreement was probably guessing and indicated no misconception (see Eitel et al., 2021). The stronger participants believed in the misconception, the more certain participants were in their agreement with a false statement (e.g., scores of 2 vs. 4 in the certainty rating). Accordingly, a stronger misconception was indicated by a higher misconception score (e.g., 2 vs. 4). Using this multiplication method, the range of possible values per item was -4 to +4, making it possible to approximate the level of interval-scaled data required to perform confirmatory factor analyses with (robust) maximum likelihood estimation (see Eitel et al., 2021).

We assumed a misconception to be prevalent, whenever participants answered at least one of the five items per misconception scale *incorrectly with high certainty* (i.e., response certainty of 3 or higher, on scale from 0 to 4; see previous section). We did so because a *mixed (mis-)conception* would be prevalent in that case (see Vosniadou, 1994). Misconceptions can be very extreme ("The earth is flat"), but they can also be "alleviated" by integrating correct information ("The earth is round"). However, this alleviation may lead to a so-called *mixed misconception* ("The earth is round, but where we stand on it, it must be flat for us not to fall off"). This would still require further refutation (Vosniadou, 1994). One incorrect answer per misconception scale (made with high certainty) already indicates such a (mixed) misconception, which requires refutation in order to achieve a correct conception (Vosniadou, 1994; see Dersch et al., 2022).

### Data Analysis

We used IBM SPSS statistics® for data preprocessing and item statistics. We used R for statistical computing (R Core Team, 2017; version 3.6.23) with the *psych* package for reliability analyses (Revelle and Condon, 2019). We calculated McDonald's omega ( $\omega$ ) for robust reliability estimation even when item-scale correlations are not tau-equivalent (Deng and Chan, 2017). Asymptotic omega simulates the theoretical omega obtained for a test of infinite length with a structure similar to the observed test. Modest reliability for McDonald's omega is at

around 0.70 (Nunnally, 1978). However, this convention should be considered with some caution as satisfactory values depend on the measurement purpose (e.g., group statistics or individual assessment) and on the nature of the scale. If assessing broad or heterogeneous constructs, even relatively low coefficients of criterion reliability (e.g., 0.50) do not seriously attenuate validity coefficients (Schmitt, 1996).

We used the *lavaan* package for confirmatory factor analysis (Rosseel, 2012) to inspect the internal structure of the MGMQ by estimating its construct validity. We used maximum likelihood estimation with robust standard errors (MLR) to handle our interval data with moderate deviations from the normal distribution (Li, 2016). We considered the global model fit to be sufficiently good if the following criteria were met: a CFI (comparative fit index) value equal to or higher than 0.95, a root mean square error of approximation (RMSEA) smaller than 0.06 (Hu and Bentler, 1998), and a standardized root mean square residual (SRMR) smaller than or equal to 0.07 (Yu, 2002). We considered the local model fit to be acceptable if values for the fully standardized factor loadings were statistically significant ( $p < 0.05$ ) and higher than 0.30 (Nunnally, 1978; Cristobal et al., 2007).

## RESULTS

### Structure Hypothesis

We first examined the MGMQ's factorial structure by comparing global and local fit measures of two structural models against each other in a confirmatory factor analysis. We expected the MGMQ data to better fit a correlated three-factor model of math-gender misconceptions (*empathizing-systemizing*, *girls' compensation*, and *girls' non-compensability*) than a general-factor model with one misconception construct. Accordingly, results revealed an overall acceptable global fit for the three-factor model (with five items per factor),  $CFI = 0.94$ ,  $RMSEA = 0.058$ ,  $SRMR = 0.057$ ,  $\chi^2 = 157.75$ ,  $df = 87$ ,  $p < 0.001$ . The factors *girls' compensation* and *non-compensability* were highly positively correlated to each other ( $r = 0.86$ ,  $p < 0.001$ ), and to *empathizing-systemizing* ( $r = 0.72$ ,  $p < 0.001$ ;  $r = 0.51$ ,  $p < 0.001$ ). Results revealed an unacceptable global fit for the general-factor model,  $CFI = 0.80$ ,  $RMSEA = 0.10$ ,  $SRMR = 0.08$ ,  $\chi^2 = 296.98$ ,  $df = 90$ ,  $p < 0.001$ . Supporting the structure hypothesis, the model fit of the three-factor model was statistically significantly better than the fit of the general-factor model,  $\chi^2(3) = 62.50$ ,  $p < 0.001$ . On the level of local model fit, factor loadings were all significant (all  $ps < 0.01$ ) and ranged between 0.44 and 0.83 for the three-factor model ( $M = 0.62$ ,  $SD = 0.12$ ; see Figure 2). Scale reliabilities [using McDonald's omega ( $\omega$ )] were good for *empathizing-systemizing* ( $\omega = 0.88$ ), acceptable for *girls' compensation* ( $\omega = 0.76$ ), and acceptable for *girls' non-compensability* ( $\omega = 0.72$ ).

### Prevalence Hypothesis

We expected student teachers to rather endorse the first two of the three gender misconceptions about mathematical abilities. As expected, more student teachers believed

that boys are inherently better in mathematics because they think more systematically (*empathizing-systemizing*; 32.0%), and that girls are only as good in mathematics as boys because they work harder (*girls' compensation*; 26.7%) and that girls cannot compensate for their lower mathematical abilities (*girls' non-compensability*; 17.5%). Overall, 14.2% of student teachers endorsed both the *empathizing-systemizing* and the *girls' compensation* misconception, whereas 44.6% of student teachers endorsed at least one of these two misconceptions. In total, 48.2% of student teachers endorsed at least one of the three misconceptions. However, on average, student teachers had negative misconception values in the MGMQ (see **Table 1**). This indicates that the majority of student teachers – correctly – disagreed with the misconception items and did not hold (strong) math-gender misconceptions.

### Association Hypothesis

We expected math-gender misconceptions to be positively associated with the prevalence of math-gender stereotypes. We found that 141 out of 303 student teachers indicated math to be more male than female, yielding a prevalence rate of 46.5%. A total of 150 student teachers (49.5%) indicated math to be equally male and female, whereas only 12 student teachers (4.0%) indicated math to be more female than male. Overall, the latent correlation between math-gender misconceptions and holding the math-gender stereotype was moderate,  $r = 0.45$ ,  $p < 0.001$ . Descriptively, we found that the *empathizing-systemizing*,  $r = 0.43$ ,  $p < 0.001$ , and the *girls' compensation* misconception,  $r = 0.44$ ,  $p < 0.001$ , correlated stronger with holding the math-gender stereotype than the *girls' non-compensability* misconception,  $r = 0.25$ ,  $p = 0.01$ .

Apart from that, we expected holding a fixed ability mindset for mathematics (Dweck, 1999; Leslie et al., 2015) to correlate more positively with the *girls' non-compensability* than with the *girls' compensation* misconception. We found that student teachers with a stronger fixed ability mindset for mathematics believed more strongly in all three misconceptions ( $r = 0.28$ ,  $p < 0.001$ ), however, not to a stronger degree in the *girls' non-compensability* misconception ( $r = 0.22$ ,  $p = 0.004$ ) than in the *girls' compensation* misconception ( $r = 0.28$ ,  $p < 0.001$ ).

## DISCUSSION

Math-gender stereotypes held by important socializers like teachers may be contributing to the underrepresentation of girls and women in STEM (for a review, see Gunderson et al., 2012). The goal of this research was to explore the *specific* misconceptions underlying math-gender stereotypes in a student teacher sample. To this end, we first analyzed the structure and prevalence of three potential misconceptions using the newly developed Math Gender Misconceptions Questionnaire (MGMQ). Afterward, we inspected to what degree holding these

misconceptions related to holding math-gender stereotypes, and fixed mindsets about math ability.

### Structure of Math-Gender Misconceptions Amongst Preservice Teachers

We constructed the MGMQ to uncover a three-factor structure of misconceptions about gender differences in mathematical abilities that we expected to observe based on prior research: *empathizing-systemizing*, *girls' compensation* and *girls' non-compensability*. We obtained evidence for the supposed three-factor structure via confirmatory factor analysis. The three-factor model fit the data better than the model assuming one general misconception factor (see **Figure 2**). Math-gender misconceptions are thus expressed through three distinct factors. (1) There is the *empathizing-systemizing* misconception assuming that pre-natal testosterone-exposure levels are lower in girls than in boys, which leads to girls thinking less systematically in relation to more empathically. Girls' less systematic thinking – according to this misconception – leads to girls' lower mathematical abilities (Baron-Cohen, 2005). (2) The *girls' compensation* misconception assume that girls are more hardworking than boys, resulting in their equally good performance in math (e.g., equal grades; Tiedemann, 2002; Sáinz et al., 2020). (3) The *girls' non-compensability* misconception assumes that girls are not only less talented in math – for example due to the *empathizing-systemizing* misconception – but furthermore, they lack the means to compensate for their disadvantage, as math talent is fixed (Dweck, 1999; Leslie et al., 2015).

The *empathizing-systemizing* scale showed good reliability; all items correlated substantially with the construct (see **Table 1**). The *girls' compensation* and - *non-compensability* scales showed acceptable reliabilities. The higher reliability of the *empathizing-systemizing* scale, compared to the other two scales, may be due to the items of *empathizing-systemizing* being very homogeneous; they all referred to the explanation of talent differences in boys and girls in mathematics. Items on the other two scales referred to both the talent differences in boys and girls in mathematics and the consequences of such talent differences. Items on the *girls' compensation* scale refer to (1) girls having less talent in mathematics, and (2) girls usually compensating for their lesser talent. Items on the *girls' non-compensability* scale refer to (1) girls having less talent in mathematics, and (2) how girls should be treated to adapt to their lack of talent (lower standards for girls; see **Table 1**, for an overview of all items). Meaning, *girls' compensation* as well as *girls' non-compensability* are broader and more heterogeneous constructs, which may explain their lower reliability coefficients than for the *empathizing-systemizing* scale.

### Prevalence and Correlates of Math-Gender Misconceptions

Almost half of the preservice teachers (48.2%) held at least one of the three misconceptions. A majority of student teachers, however, held no math-gender misconceptions, even according to the strict criteria we applied. This finding led to negative

average math-gender misconception scores among student teachers in this sample (see **Table 1**), which imply that on average, math-gender misconceptions are not (strongly) prevalent. These results are encouraging, even if they are still far from ideal. The prevalence of math-gender misconceptions among a subgroup of student teachers is still worrying, since even endorsing just one the misconceptions can affect teachers' instruction. As a consequence, misconceptions may cause a different treatment of the genders (e.g., Carlana, 2019), and reinforce math-gender stereotypes among schoolchildren (e.g., Geis, 1993; Eccles, 2011; Gunderson et al., 2012). The math-gender stereotypes weaken female representation in mathematical careers (e.g., Eccles, 2011; Wang and Degol, 2017). As teachers function as multipliers of their own knowledge and beliefs and teach many students during their career, misconceptions deserve attention and interventions in teacher education and training, even if only a subgroup of teachers seems to endorse such misconceptions.

As expected, both the *empathizing-systemizing* (32.0%) and *girls' compensation* misconception (26.7%) seemed to be more prevalent than the *girls' non-compensability* (17.5%) misconception. This difference in prevalence may partially be due to social desirability. Agreeing to the *empathizing-systemizing* misconception may be more socially desirable than agreeing to statements on the two other misconception scales, because the former statements (1) highlight girls' empathic and social abilities and (2) provide an explanation for girls' lack of talent that did not blame the girls themselves, but rather their genes or pre-natal influences on their body. Like for benevolent sexism (Glick and Fiske, 1996, 1997) these two apparently "positive" beliefs about girls might have been more socially acceptable than agreeing with the beliefs captured by the other two misconceptions.

The other two misconceptions consisted of statements displaying obvious, less benevolent sexism, such as indicating that (1) girls lack talent and (2) the genders should be treated differently and thus unequally. Such attitudes tend to be rejected nowadays among well-educated students in Western societies, like those in our study sample: Accordingly, the students in our sample indicated moderate to high agreement with feminism, which correlated negatively with misconception endorsement ( $r = -0.21$ ,  $p = 0.001$ ). This lower social desirability thus may have reduced agreement rates with the *girls' compensation* and *girls' non-compensability* scale, even though actual beliefs may differ from what participants indicated. Furthermore, the awareness that fixed mindsets in teachers are detrimental to their students (e.g., Canning et al., 2019, 2021; Heyder et al., 2020) seems to be increasing in (teacher) education (Dweck, 2016). Thus, especially the *girls' non-compensability* scale – theoretically a combination of fixed ability mindset ideas and promoting girls' lesser abilities, might be perceived as socially undesirable, which could have contributed to the (relatively speaking), lowest endorsement rates.

Furthermore, the significant correlation between fixed mindset in math and the *girls' non-compensability* misconception as well as the non-significant correlation with the *girls' compensation* misconception supports the construct validity of the MGMQ's constructs: It is only when abilities are perceived as fixed that there is no way to compensate for low abilities.

Since fixed mindsets in math have been found to be detrimental only in terms of female students' intrinsic motivation and ability self-concepts (Heyder et al., 2020), these associations further corroborate the importance of the *girls' non-compensability* misconception for female students' engagement in math.

Our findings also support prior research findings of (preservice) teachers holding explicit math-gender stereotypes (e.g., Li, 1999; Tiedemann, 2002; Cimpian et al., 2016; Sáinz et al., 2020). Also in our study, about half (49.5%) of the preservice teachers held explicit math-gender stereotypes. These explicit math-gender stereotypes were associated with math-gender-misconceptions to a moderate degree ( $r = 0.45$ ), tentatively supporting the idea of math-gender-misconceptions underlying math-gender stereotypes. So far, math-gender stereotypes have been assessed either via implicit association testing (e.g., Nosek et al., 2010; Steffens and Jelenec, 2011), or ratings of whether math is more female than male (e.g., Nosek et al., 2010), or via one to three simple items about talent differences (see Hyde et al., 1990; Gunderson et al., 2012). The current assessment of math-gender misconceptions as a construct underlying math-gender stereotypes is a novel approach to understand and potentially refute math-gender gender stereotypes. In the future, assessing math-gender misconceptions in addition to math-gender stereotypes may facilitate the comprehension of math-gender stereotypes and thus our ability to target both – math-gender stereotypes and math-gender misconceptions. With this reasoning, it is important to note that holding the math-gender stereotype correlated most strongly with holding the *empathizing-systemizing* and the *girls' compensation* misconception. Specifically targeting these misconceptions (e.g., by means of refutation text; Tippett, 2010) may thus be a promising means to reduce not just the specific misconception but also math-gender stereotypes to a certain degree. More research applying more measures for math-gender stereotypes and evaluating their association with math-gender misconceptions is necessary to gain more insights into the association between math-gender misconceptions, implicit and explicit math-gender stereotypes, as well as how they manifest in teacher and student teacher behavior. Additionally, applying more measures of explicit math-gender stereotypes in future research to assess the relations between math-gender stereotypes and math-gender misconceptions should help further validate the MGMQ in future research. Hence, this study is the first of a planned series of studies on the relationship between math-gender misconceptions and math-gender stereotypes.

## Limitations and Further Research

In this study, we presented the MGMQ, a novel measure assessing misconceptions about gender differences in math abilities. To the best of our knowledge, this is the first study applying the concept of misconceptions (e.g., Eitel et al., 2019) to the important field of women's underrepresentation in math. Therefore, some limitations and questions for future research emerged.

First of all, as the main objective of this research was the construction and evaluation of the MGMQ, we implemented only *one* measure to assess math-gender stereotypes [similarly applied by Nosek et al. (2010)]. It is certainly useful to relate

the MGMQ results to other measures assessing math-gender stereotypes in further research. In this paper, we described the MGMQ development. As the MGMQ has demonstrated its reliability as a measuring tool within our sample, we intend to further research its reliability and interrelations between the MGMQ scales and various implicit and explicit – as well as behavioral stereotype-measures in future research.

Another limitation refers to the risk of triggering socially desirable responses as discussed before. Furthermore, recognizing and reporting socially undesirable stereotypes may require a certain degree of self-awareness among participants (Nosek, 2007). Some may not have thought about their stereotypes because they were unwilling to. But even though reflection is necessary and social desirability may hinder the readiness to self-report stereotypes, direct self-reporting is still known to work best for assessing stereotypes (Axt, 2018). In future research, some items (e.g., “Since girls are on average less mathematically gifted, they should be assessed with different criteria than boys”) could be revised to make them more neutral-sounding. Strongly overlapping items could be excluded, forming a short version of the MGMQ (e.g., “As girls think more empathically whereas boys think more systematically, boys are on average more talented for math than girls”). A short version should be economic and especially practical for applying it to in-service teachers, as they have less time to participate in research. In future studies, it would be also promising for researchers to stress that the MGMQ is a knowledge test, not an attitude test, thus hopefully reducing further answer bias due to social desirability. Future research with the MGMQ could also focus on the prevalence of math-gender misconceptions in math teachers, as math teachers, due to their direct influence on girls’ math learning, may contribute especially to the upholding of math-gender misconceptions (and math-gender stereotypes). In this regard, we compared the misconception prevalence between student teachers with and without mathematics as teaching subjects here. We observed small and insignificant differences between students with math ( $M = -1.64$ ,  $SD = 1.17$ ) and without math as teaching subject ( $M = -1.48$ ,  $SD = 1.40$ ),  $t(301) = 1.04$ ,  $p = 0.30$ .

Further, our sample’s gender distribution consisting of 79.9% women does not represent the general population. However, this high percentage of women in our student teacher sample resembles the gender distribution of teachers in Germany: The Federal Office for Statistics in Germany assessed teachers’ gender in the school year of 2019/2020 and found that 73.1% of teachers in general education were female. As gender might still have influenced the math-gender misconception prevalence, we compared the prevalence rates between genders, and revealed that the prevalence of math-gender misconceptions did not differ between female participants ( $M = -1.56$ ,  $SD = 1.33$ ) and male participants ( $M = -1.46$ ,  $SD = 1.27$ ),  $t(301) = 0.54$ ,  $p = 0.58$ . This insignificant difference may be due to the exposure to math-gender misconceptions in our society regardless of gender.

Furthermore, implicit and behavioral measures could support the assessment and generate additional knowledge about the prevalence of math-gender misconceptions or math-gender stereotypes.

The goal of the MGMQ is to identify math-gender misconceptions that potentially underlie math-gender stereotypes. As (math-gender) stereotypes have rarely been successfully reduced (FitzGerald et al., 2019; Kollmayer et al., 2020), identifying underlying math-gender misconceptions is a starting point for conceptual change – and hopefully attitude change as well. Interventions targeting misconceptions among teachers (e.g., refutation texts; Menz et al., 2021) could therefore also be applied to revise or reduce stereotypes among teachers.

In addition to the math-gender misconceptions discussed here, there are misconceptions and ideas associated with other stereotypes that influence math representation and warrant research. This should yield insights on whether such associations between stereotypes and misconceptions are specific to the gender topic, or generalizable. One example would be math-race stereotypes (Starr and Simpkins, 2021). The intersectionality of stereotypes, meaning people belonging to more than one minority group (e.g., Black and female) and thus suffering from different overlapping adverse stereotypes, should be considered in future research (Yuval-Davis, 2006; Parker et al., 2020).

## Conclusion

This study describes a newly developed instrument assessing misconceptions about gender differences in math ability that potentially underlie gender stereotypes, and which therefore may contribute to the underrepresentation of women in math careers. Our results show that (a) our newly developed questionnaire reliably assessed three distinct misconceptions related to gender differences in mathematics in the first sample, (b) almost half of the participating preservice teachers endorsed at least one of the three misconceptions, whereas a majority did not, and (c) holding these misconceptions was substantially associated with holding math-gender stereotypes.

Identifying the specific misconceptions potentially behind math-gender stereotypes is a good starting point for interventions aiming at conceptual change (Larkin, 2012), also in the field of gender and STEM. Since misconceptions hinder the acquisition of scientifically accurate conceptions (Eitel et al., 2021), overcoming them is important to reduce gender disparities in STEM in the future. This study provides the basis upon which to develop specific instructions in the form of refutation texts during teacher education or training (Eitel et al., 2019; Menz et al., 2021; Dersch et al., 2022).

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Lokale Ethik-Kommission des

Fachbereichs 06 der Justus-Liebig-Universität Gießen. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

A-SD, AH, and AE contributed to the conception and design of the study and wrote the sections of the manuscript. A-SD organized the database and wrote the first draft of the manuscript.

A-SD and AE performed the statistical analysis. All authors contributed to manuscript revision, read, and approved the submitted version.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.820254/full#supplementary-material>

## REFERENCES

- Avitzour, E., Choen, A., Joel, D., and Lavy, V. (2020). *On the Origins of Gender-Biased Behavior: The Role of Explicit and Implicit Stereotypes* (No. w27818). Cambridge, MA: National Bureau of Economic Research, doi: 10.3386/w27818
- Axt, J. R. (2018). The best way to measure explicit racial attitudes is to ask about them. *Soc. Psychol. Personal. Sci.* 9, 896–906. doi: 10.1177/1948550617728995
- Baron-Cohen, S. (2005). The essential difference: the male and female brain. *Phi Kappa Phi Forum* 85, 23–26.
- Bian, L., Leslie, S. J., and Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science* 355, 389–391. doi: 10.1126/science.aah6524
- Bloom, H. S., Hill, C. J., Black, A. B., and Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *J. Res. Educ. Effect.* 1, 289–328. doi: 10.1080/19345740802400072
- Canning, E. A., Muenks, K., Green, D. J., and Murphy, M. C. (2019). STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Sci. Adv.* 5:eaa4734. doi: 10.1126/sciadv.aau4734
- Canning, E. A., Ozier, E., Williams, H. E., AlRasheed, R., and Murphy, M. C. (2021). Professors Who Signal a Fixed Mindset About Ability Undermine Women's Performance in STEM. *Soc. Psychol. Personal. Sci.* 2021:19485506211030398. doi: 10.1177/19485506211030398
- Carlana, M. (2019). Implicit Stereotypes: evidence from Teachers' Gender Bias. *Q. J. Econ.* 134, 1163–1224. doi: 10.1093/qje/qjz008
- Chi, M. T., and Roscoe, R. D. (2002). The processes and challenges of conceptual change. *Iss. Theory Pract.* 2002, 3–27. doi: 10.1007/0-306-47637-1\_1
- Cimpian, J. R., Lubienski, S. T., Timmer, J. D., Makowski, M. B., and Miller, E. K. (2016). Have gender gaps in math closed? Achievement, teacher perceptions, and learning behaviors across two ECLS-K cohorts. *AERA Open* 2:23328588416673617. doi: 10.1177/2332858416673617
- Cristobal, E., Flavian, C., and Guinaliu, M. (2007). Perceived e-service quality (PeSQ): measurement validation and effects on consumer satisfaction and web site loyalty. *Manag. Serv. Q.* 17, 317–340. doi: 10.1108/09604520710744326
- Deng, L., and Chan, W. (2017). Testing the difference between reliability coefficients alpha and omega. *Educ. Psychol. Meas.* 77, 185–203. doi: 10.1177/0013164416658325
- Dersch, A., Renkl, A., and Eitel, A. (2022). Personalized refutation texts best stimulate teachers' conceptual change about multimedia learning. *J. Comput. Assist. Learn.* doi: 10.1111/jcal.12671
- Dweck, C. (1999). *Self-Theories: Their role in motivation, personality and development*. Hove: Taylor and Francis/Psychology Press.
- Dweck, C. (2015). Carol Dweck revisits the growth mindset. *Educ. Week* 35, 20–24. doi: 10.1111/bjep.12072
- Dweck, C. (2016). What having a "growth mindset" actually means. *Harv. Bus. Rev.* 13, 213–226.
- Eccles, J. (2011). Gendered educational and occupational choices: applying the Eccles et al. model of achievement-related choices. *Internat. J. Behav. Dev.* 35, 195–201. doi: 10.1177/0165025411398185
- Eccles, J., Adler, T. F., Futtermann, R., Goff, S. B., Kaczala, C. M., Meece, J. L., et al. (1983). "Expectations, values and academic behaviors," in *Perspective on achievement and achievement motivation*, ed. J. T. Spence (Cambridge, MA: Academic Press), 75–146.
- Eitel, A., Dersch, A.-S., and Renkl, A. (2019). Wissenschaftliche Annahmen und Fehlkonzepte angehender Lehrkräfte über das Lernen mit Multimedia. *Unterrichtswissenschaft* 47, 451–474. doi: 10.1007/s42010-019-00049-4
- Eitel, A., Prinz, A., Kollmer, J., Niessen, L., Russow, J., Ludäscher, M., et al. (2021). The misconceptions about multimedia learning questionnaire. An empirical evaluation study with teachers and student teachers. *Psychol. Learn. Teach.* 20, 420–444. doi: 10.1177/14757257211028723
- Eliot, L. (2010). *Pink brain, blue brain: How small differences grow into troublesome gaps-and what we can do about it*. New York, NY: Simon and Schuster, doi: 10.14507/er.v0.1310
- Else-Quest, N. M., Hyde, J. S., and Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychol. Bull.* 136, 103–127. doi: 10.1037/a0018053
- Escovar, E., Rosenberg-Lee, M., Uddin, L. Q., and Menon, V. (2016). The empathizing-systemizing theory, social abilities, and mathematical achievement in children. *Sci. Rep.* 6, 1–11. doi: 10.1038/srep23011
- FitzGerald, C., Martin, A., Berner, D., and Hurst, S. (2019). Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: a systematic review. *BMC Psychol.* 7, 1–12. doi: 10.1186/s40359-019-0299-7
- Geis, F. L. (1993). "Self-fulfilling prophecies: A social psychological view of gender," in *The psychology of gender*, eds A. E. Beall and R. J. Sternberg (New York, NY: Guilford Press), 9–54.
- Glick, P., and Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *J. Personal. Soc. Psychol.* 70:491. doi: 10.1037/0022-3514.70.3.491
- Glick, P., and Fiske, S. T. (1997). Hostile and benevolent sexism: Measuring ambivalent sexist attitudes toward women. *Psychol. Women Q.* 21, 119–135. doi: 10.1111/j.1471-6402.1997.tb00104.x
- Gottfredson, L. S. (1998). The general intelligence factor. *Sci. Am.* 9, 24–29.
- Greenberg, D. M., Warrier, V., Allison, C., and Baron-Cohen, S. (2018). Testing the Empathizing-Systemizing theory of sex differences and the Extreme Male Brain theory of autism in half a million people. *Proc. Natl. Acad. Sci.* 115, 12152–12157. doi: 10.1073/pnas.1811032115
- Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., and Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychol. Rev.* 109:3. doi: 10.1037/0033-295X.109.1.3
- Gunderson, E. A., Hamdan, N., Sorhagen, N. S., and D'Estes, A. P. (2017). Who needs innate ability to succeed in math and literacy? Academic-domain-specific theories of intelligence about peers versus adults. *Dev. Psychol.* 53, 1188–1205. doi: 10.1037/dev0000282
- Gunderson, E. A., Ramirez, G., Levine, S. C., and Beilock, S. L. (2012). The Role of Parents and Teachers in the Development of Gender-Related Math Attitudes. *Sex Roles* 66, 153–166. doi: 10.1007/s11199-011-9996-2
- Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., and Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychol. Sci. Public Inter.* 8, 1–51. doi: 10.1111/j.1529-1006.2007.00032.x
- Hand, S., Rice, L., and Greenlee, E. (2017). Exploring teachers' and students' gender role bias and students' confidence in STEM fields. *Soc. Psychol. Educ.* 20, 929–945. doi: 10.1007/s11218-017-9408-8
- Hannover, B. (2008). "Vom biologischen zum psychologischen Geschlecht: Die Entwicklung von Geschlechtsunterschieden," in *Lehrbuch Pädagogische Psychologie*, ed. A. Renkl (Edison: Huber), 339–388.

- Heyder, A., Steinmayr, R., and Kessels, U. (2019). Do teachers' beliefs about math aptitude and brilliance explain gender differences in children's math ability self-concept? *Front. Educ.* 4, 1–11. doi: 10.3389/educ.2019.00034
- Heyder, A., Weidinger, A. F., Cimpian, A., and Steinmayr, R. (2020). Teachers' belief that math requires innate ability predicts lower intrinsic motivation among low-achieving students. *Learn. Instruct.* 65:101220. doi: 10.1016/j.learninstruc.2019.101220
- Holder, K., and Kessels, U. (2017). Gender and ethnic stereotypes in student teachers' judgments: a new look from a shifting standards perspective. *Soc. Psychol. Educ.* 20, 471–490. doi: 10.1007/s11218-017-9384-z
- Hu, L. T., and Bentler, P. M. (1998). Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychol. Methods* 3, 424–453. doi: 10.1037/1082-989X.3.4.424
- Hughes, S., Lyddy, F., and Lambe, S. (2013). Misconceptions about Psychological Science: a Review. *Psychol. Learn. Teach.* 12, 20–31. doi: 10.2304/plat.2013.12.1.20
- Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., and Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: a meta-analysis. *Psychol. Wom. Q.* 14, 299–324. doi: 10.1111/j.1471-6402.1990.tb00022.x
- Keller, C. (2001). Effect of Teachers' Stereotyping on Students' Stereotyping of Mathematics as a Male Domain. *J. Soc. Psychol.* 141, 165–173. doi: 10.1080/00224540109600544
- Kersey, A. J., Csusmitta, K. D., and Cantlon, J. F. (2019). Gender similarities in the brain during mathematics development. *NPJ Sci. Learn.* 4, 1–7. doi: 10.1038/s41539-019-0057-x
- Klineberg, O. (1951). The scientific study of national stereotypes. *Internat. Soc. Sci. Bull.* 3, 505–514.
- Kollmayer, M., Schultes, M. T., Lüftenegger, M., Finsterwald, M., Spiel, C., and Schober, B. (2020). REFLECT-a teacher training program to promote gender equality in schools. *Front. Educ.* 5:136. doi: 10.3389/educ.2020.00136
- Lahance, J. A., and Mazzocco, M. M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learn. Indiv. Diff.* 16, 195–2016. doi: 10.1016/j.lindif.2014.03.020
- Larkin, D. (2012). Misconceptions about "misconceptions": Preservice secondary science teachers' views on the value and role of student ideas. *Sci. Educ.* 96, 927–959. doi: 10.1002/sce.21022
- Leslie, S. -J., Cimpian, A., Meyer, M., and Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science* 347, 262–265. doi: 10.1126/science.1261375
- Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behav. Res. Methods* 48, 936–949. doi: 10.3758/s13428-015-0619-7
- Li, Q. (1999). Teachers' beliefs and gender differences in mathematics: a review. *Educ. Res.* 41, 63–76. doi: 10.1080/0013188990410106
- Menz, C., Spinath, B., Hendriks, F., and Seifried, E. (2021). Reducing educational psychological misconceptions: how effective are standard lectures, refutation lectures, and instruction in information evaluation strategies? *Scholar. Teach. Learn. Psychol.* 2021:269. doi: 10.1037/stl0000269
- Merton, R. K., and Merton, R. C. (1968). *Social theory and social structure*. New York, NY: Simon and Schuster.
- Miller, D. I., Eagly, A. H., and Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations. *J. Educ. Psychol.* 107:631. doi: 10.1037/edu0000005
- Mischel, W. (1966). "A social-learning view of sex differences in behavior," in *The development of sex differences*, ed. E. E. Maccoby (Redwood City, CA: Stanford University Press), 57–81.
- Moosbrugger, H., and Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion*. New York, NY: Springer. doi: 10.1007/978-3-642-20072-4\_1
- Muenks, K., Canning, E. A., LaCasse, J., Green, D. J., Zirkel, S., Garcia, J. A., et al. (2020). Does my professor think my ability can change? Students' perceptions of their STEM professors' mindset beliefs predict their psychological vulnerability, engagement, and performance in class. *J. Exp. Psychol. Gen.* 149, 2119–2144. doi: 10.1037/xge0000763
- Muntoni, F., and Retelsdorf, J. (2018). Gender-specific teacher expectations in reading—the role of teachers' gender stereotypes. *Contemp. Educ. Psychol.* 54, 212–220. doi: 10.1016/j.cedpsych.2018.06.012
- Nosek, B. A. (2007). Implicit–explicit relations. *Curr. Direct. Psychol. Sci.* 16, 65–69. doi: 10.1111/j.1467-8721.2007.00477.x
- Nosek, B. A., and Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *Am. Educ. Res. J.* 48, 1125–1156. doi: 10.3102/0002831211410683
- Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Banaji, M. R., et al. (2010). Pervasiveness and correlates of implicit attitudes and stereotypes. *Eur. Rev. Soc. Psychol.* 18, 36–88. doi: 10.1080/10463280701489053
- Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York, NY: McGraw-Hill.
- Parker, P. D., Van Zanden, B., Marsh, H. W., Owen, K., Duineveld, J. J., and Noetel, M. (2020). The intersection of gender, social class, and cultural context: a meta-analysis. *Educ. Psychol. Rev.* 32, 197–228. doi: 10.1007/s10648-019-09493-1
- Passolunghi, M. C., Rueda Ferreira, T. I., and Tomasello, C. (2014). Math-gender stereotypes and math-related beliefs in childhood and early adolescence. *Learn. Indiv. Diff.* 34, 70–76. doi: 10.1016/j.lindif.2014.05.005
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Reilly, D. (2012). Gender, culture and sex-typed cognitive abilities. *PLoS One* 7:e39904. doi: 10.1371/journal.pone.0039904
- Reilly, D., Neumann, D. L., and Andrews, G. (2015). Sex differences in mathematics and science achievement: a meta-analysis of National Assessment of Educational Progress assessments. *J. Educ. Psychol.* 107, 645–662. doi: 10.1037/edu0000012
- Retelsdorf, J., Schwartz, K., and Asbrock, F. (2015). "Michael can't read!" – Teachers' gender stereotypes and boys' reading self-concept. *J. Educ. Psychol.* 107, 186–194. doi: 10.1037/a0037107
- Revelle, W., and Condon, D. M. (2019). Reliability from  $\alpha$  to  $\omega$ : a tutorial. *Psychol. Assess.* 31:1395. doi: 10.1037/pas0000754
- Riegle-Crumb, C., and Humphries, M. (2012). Exploring Bias in Math Teachers' Perceptions of Students' Ability by Gender and Race/Ethnicity. *Gen. Soc.* 26:2. doi: 10.1177/0891243211434614
- Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., and Copur-Gencturk, Y. (2014). Teachers' perceptions of students' mathematics proficiency may exacerbate early gender gaps in achievement. *Dev. Psychol.* 50, 1262–1281. doi: 10.1037/a0035073
- Rodriguez, S., Regueiro, B., Piñeiro, I., Estévez, I., and Valle, A. (2020). Gender differences in mathematics motivation: differential effects on performance in primary education. *Front. Psychol.* 10:3050. doi: 10.3389/fpsyg.2019.03050
- Rossee, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *J. Stat. Softw.* 48, 1–36. doi: 10.1002/9781119579038.ch1
- Sáinz, M., Fábregues, S., and Solé, J. (2020). Parent and teacher depictions of gender gaps in secondary student appraisals of their academic competences. *Front. Psychol.* 11:573752. doi: 10.3389/fpsyg.2020.573752
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychol. Assess.* 8, 350–353. doi: 10.1037/1040-3590.8.4.350
- Skaalvik, E. M., and Rankin, R. J. (1994). Gender differences in mathematics and verbal achievement, self-perception and motivation. *Br. J. Educ. Psychol.* 64, 419–428. doi: 10.1111/j.2044-8279.1994.tb01113.x
- Skaalvik, E. M., Federici, R. A., and Klassen, R. M. (2015). Mathematics achievement and self-efficacy: Relations with motivation for mathematics. *Internat. J. Educ. Res.* 72, 129–136. doi: 10.1016/j.ijer.2015.06.008
- Smith, A. E., Jussim, L., and Eccles, J. (1999). Do self-fulfilling prophecies accumulate, dissipate, or remain stable over time? *J. Personal. Soc. Psychol.* 77:548. doi: 10.1037/0022-3514.77.3.548
- Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and women's math performance. *J. Exp. Soc. Psychol.* 35, 4–28. doi: 10.1006/jesp.1998.1373
- Starr, C. R., and Simpkins, S. D. (2021). High school students' math and science gender stereotypes: relations with their STEM outcomes and socializers' stereotypes. *Soc. Psychol. Educ.* 24, 273–298. doi: 10.1007/s11218-021-09611-4
- Steele, C. M., and Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *J. Personal. Soc. Psychol.* 69:797. doi: 10.1037/0022-3514.69.5.797
- Steffens, M. C., and 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, 324–335. doi: 10.1007/s11199-010-9924-x

- Tiedemann, J. (2002). Teachers' gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educ. Stud. Math.* 50, 49–62. doi: 10.1023/A:1020518104346
- Tippett, C. D. (2010). Refutation text in science education: a review of two decades of research. *Internat. J. Sci. Educ.* 8, 951–970. doi: 10.1007/s10763-010-9203-x
- Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and instruction* 4, 45–69. doi: 10.1016/0959-4752(94)90018-3
- Wang, M.-T., and Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): current knowledge, implications for practice, policy, and future directions. *Educ. Psychol. Rev.* 29, 119–140. doi: 10.1007/s10648-015-9355-x
- Wentzel, K. R. (2014). "Socialization in school settings," in *Handbook of socialization: Theory and research, Second Edition*, eds J. E. Grusec and P. D. Hastings (New York, NY: The Guilford Press), 251–275.
- Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* [Unpublished doctoral dissertation]. Los Angeles, CA: University of California.
- Yuval-Davis, N. (2006). Intersectionality and feminist politics. *Eur. J. Women Stud.* 13, 193–209. doi: 10.1177/1350506806065752
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Dersch, Heyder and Eitel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

-This manuscript is currently under Review at Social Psychology of Education-

**Teachers' Awareness of Math-gender Stereotype-reinforcing Behaviors in the Classroom:  
A Video Experiment**

Anna-Sophia Dersch<sup>1</sup>, <https://orcid.org/0000-0002-6012-3411>

Anke Heyder<sup>2</sup>, <https://orcid.org/0000-0002-1578-4894>

Alexander Eitel<sup>1</sup>, <https://orcid.org/0000-0001-6403-3467>

<sup>1</sup> Department of Educational Psychology, University of Giessen, Germany

<sup>2</sup> Department of Educational Psychology, University of Bochum, Germany

Corresponding Author: Anna-Sophia Dersch

Justus-Liebig-Universität Gießen

FB06 Lehren und Lernen mit Medien

Otto-Behaghel-Straße 10 F

35394 Gießen

Mail: [anna-sophia.dersch@psychol.uni-giessen.de](mailto:anna-sophia.dersch@psychol.uni-giessen.de)

### Abstract

The stereotype of math being male-typed prevails, also among teachers. This math-gender stereotype is related to math-gender misconceptions, which are scientifically incorrect specific theories about girls' math abilities (e.g., girls inherently think less systematically and hence have lower math abilities than boys). Teachers can contribute to the manifestation of the math-gender stereotype via their implicit or explicit behaviors (e.g., the selection of stereotyped learning materials).

Here, we researched teachers' awareness of math-gender stereotype-reinforcing behaviors in scripted classroom videos. In an online experiment, we randomly assigned 278 participating teachers to videos with a female vs. male model teacher's math-gender stereotype-reinforcing behaviors. We instructed participants to evaluate the model teacher's teaching behaviors.

Few teachers were aware of implicit math-gender stereotype-reinforcing behaviors; most were aware of explicit math-gender stereotype-reinforcing behaviors. Teachers were more aware of the male model teacher's than of the female model teacher's math-gender stereotype-reinforcing behaviors. If model gender and observer gender aligned, this improved awareness. Awareness of math-gender stereotype-reinforcing behaviors was not related to the math-gender stereotype. Awareness was related to math-gender misconceptions: Participants who agreed less with scientifically incorrect theories about girls' lower inherent math abilities were more aware of math-gender stereotype-reinforcing behaviors.

The findings of low awareness of implicit math-gender stereotype-reinforcing behaviors call for teacher training to improve teachers' awareness in the classroom.

|

*Keywords:* teacher education; classroom videos; math-gender stereotype; teacher noticing; psychology misconceptions

### Introduction

Envision a math lesson in 7<sup>th</sup> grade. The students practice fractional arithmetic. The teacher stands before the class with the open textbook and says, "This task about Mr. Meier, the road engineer, is quite good. Let's work with this one. He needs 14 tons of tar and 300 kilograms of concrete for his construction site. How many tons of concrete does he need, then? The result as a fraction, please ..." The teacher points in Thomas' direction: "Yes, Thomas, please." "  $\frac{3}{10}$  tons of concrete," Thomas answers. "That is correct," the teacher states. "I've thought of a similar task for you: Mrs. Schäfer goes to the market. For her cake recipe, she needs  $\frac{1}{3}$  kilograms of apples and  $\frac{1}{2}$  kilograms of flour. How many grams does she need of each?"

Tasks like this commonly appear in learning materials such as math textbooks. Learning materials have more male than female protagonists, and male protagonists are more often portrayed via their mathematical or STEM profession (Jehle et al., 2024; Kerkhoven et al., 2016; Moser & Hannover, 2014; Nandi et al., 2024). Such portrayal of the genders potentially reinforces the common math-gender stereotype of math being male-typed (Moser & Hannover, 2014). Most teachers are neither responsible for textbook content nor for selecting textbooks for their students. However, teachers should be more aware of these representational issues. In that case, they might be more likely to select textbook tasks that do not reinforce the math-gender stereotype among students in the way described above (see also Duguid & Thomas-Hunt, 2015).

The math-gender stereotype often contains incorrect theories about inherent ability differences between the genders, meaning *math-gender misconceptions* (Dersch et al., 2022a). Such beliefs are detrimental to female representation in mathematics and other STEM careers (Eccles, 2011; Girelli, 2023; Gunderson et al., 2012; Hand et al., 2017; Wang & Degol, 2017). If asked directly, nowadays, teachers report only a few explicit math-gender misconceptions or math-gender stereotypes (e.g., Dersch et al., 2022a). However, does this mean that teachers are aware of classroom behaviors like the task selection above that more or less explicitly reinforce the math-gender stereotype? This is the focus of the present research.

In prior research, classroom videos have been used to assess teachers' awareness of functional teaching strategies (Kersting et al., 2012). Similarly, we used scripted classroom videos of a female vs. male model teacher expressing *math-gender stereotype-reinforcing behaviors* (e.g. selecting stereotypical textbook tasks, like in the top example) to assess (1) awareness of math-gender stereotype-reinforcing behaviors, (2) the role of participant and model teacher gender regarding awareness of such behaviors and (3) the relation of awareness of math-gender stereotype-reinforcing behaviors to math-gender stereotype and math-gender misconceptions.

### The Math-gender Stereotype and Math-gender Misconceptions

Stereotypes are part of attitudes and are – to some extent collectively shared – beliefs about people based on their affiliation with particular social groups (Greenwald et al., 2002). Stereotypes are a type of cognitive schema (Fiske, 1982). Stereotyping can help structure perception and cognitive offloading (Fiske, 1982). However, stereotypes are sets of rigid, over-simplified, and over-generalized beliefs that influence the perception of those holding the stereotype and, hence, influence the social group affected by the stereotype (Ashmore & Del Boca, 1979; Klineberg, 1951). A prominent example is the *math-gender* stereotype, which is a set of over-simplified or overgeneralized beliefs of girls having weaker mathematical abilities than boys because of their gender (Dersch et al., 2022a), is prevalent among most Western societies (Nosek & Smyth, 2011; Wang & Degol, 2017). Stereotypes about math and other STEM subjects prevail and predict low female representation in said areas (Hand et al., 2017; Nosek & Smyth, 2011). The math-gender stereotype itself contributes to this low female representation in mathematics and STEM, for example, by negatively influencing female learners' identification with math (Cvencek et al., 2012; Greenwald et al., 2002) or by reducing subjective task value and expectancy of success in math (Eccles, 2011). These mechanisms then contribute to female learners being less likely to make achievement-related choices in favor of mathematics (e.g., Eccles, 2011; Eccles & Wigfield, 2024; Wang & Degol, 2017).

Stereotypes, according to most definitions, are *sets of beliefs* (Ashmore & Del Boca, 1979). Apart from reflecting the social reality of lower female representation in mathematics (e.g., Jussim et al., 2005; Kahn & Ginther, 2017), these beliefs can also contain specific, scientifically incorrect, or empirically unsupported knowledge theories, that is, *misconceptions*. Misconceptions are subjective theories incompatible with empirical evidence while appearing highly plausible (Hughes et al., 2013). Prior research has found particular support for the presence of two math-gender misconceptions of (1) *empathizing-systemizing* and (2) *girls' compensation* (Dersch et al., 2022a).

(1) According to *empathizing-systemizing*, higher prenatal testosterone exposure in boys fosters their systematic thinking in opposition to girls' rather empathic thinking. According to this misconception, the higher systemizing-to-empathizing ratio in boys leads to more inherent math talent (e.g., Baron-Cohen, 2005). First, this misconception assumes empathetic and systematic thinking to be dualities when, in fact, both relate to general intelligence (Gottfredson, 1998; Martingano & Konrath, 2022). Second, this misconception assumes an influence of systematic thinking abilities on mathematical talent. Such an influence could, however, not be proven (Escovar et al., 2016).

(2) The second misconception of *girls' compensation* assumes that girls are more diligent than boys and that this diligence leads to girls' good grades in math. In contrast, boys achieve similar grades with less effort

because they are more talented in math. While research finds girls to be more diligent than boys in math, this diligence leads to better grades (Duckworth & Seligman, 2006). Furthermore, since girls do *not* inherently lack mathematical talent (Girelli, 2023; Kersey et al., 2019; Lachance & Mazzoco, 2006; Lindberg et al., 2010), compensation for lacking mathematical talent is inconclusive.

In a prior study, these math-gender misconceptions were prevalent among pre-service teachers and correlated moderately positively with the explicit math-gender stereotype (Dersch et al., 2022a). Findings suggest that math-gender misconceptions are closely related to the math-gender stereotype but a distinct construct. Various research has furthermore provided evidence for different manifestations of the math-gender stereotype among pre-service and in-service teachers (Authors, submitted; Avitzour et al., 2020; Cimpian et al., 2016; Dersch et al., 2022a; Gunderson et al., 2012; Holder & Kessels, 2017; Rieggle-Crumb & Humpries, 2012; Robinson-Cimpian et al., 2014; for a literature review, see Li, 1999). Furthermore, various studies show relations between teachers' and students' math-gender stereotype and (e.g., Gunderson et al., 2012; Beilock et al., 2010; Keller, 2001; Starr & Simpkins, 2021). These relations might be due to teachers' math-gender stereotype-reinforcing behaviors.

### **Math-gender Stereotype-reinforcing Behaviors in Teachers**

To our knowledge, there is little research on teachers' exact math-gender stereotype-reinforcing behaviors and to what extent they are expressed since such behaviors are unlikely to be reported in self-report measures due to low social desirability of behaving math-gender stereotype-reinforcing (Agut et al., 2023; van de Mortel, 2008), and unlikely to be shown in observational settings due to increased self-awareness in such settings (e.g., Gartmeier, 2014). There are, however, several assumptions about how the math-gender stereotype may be conveyed from teachers to their students, for instance, via implicit learning (Beilock et al., 2010), model learning (Bandura, 1977a; 1977b; 1986), or operant conditioning (Mischel, 1966). Furthermore, most female students report confrontations with the math-gender stereotype (Becker & Hall, 2024; Grossman & Porche, 2014) – to a substantial extent by teachers (Leaper & Brown, 2008; Kurtz-Costes et al., 2014). Teachers can reinforce the math-gender stereotype implicitly, for instance, by selecting gender-biased learning materials, which are still commonly found in math textbooks (Jehle et al., 2024; Kerkhoven et al., 2016; Nandi et al., 2024). Representations of textbook protagonists function as role models (Moser & Hannover, 2014). Therefore, textbook tasks, which portray mostly men in math-related jobs, lead to girls lacking gender-similar role models. As model-observer-similarity increases the observers' self-efficacy (Bandura, 1977a; 1986), the lack of gender-similar models for girls in mathematics or other STEM areas may lead to girls being less likely to perceive these areas as potential careers, which contributes to the low female representation in mathematics.

Furthermore, differences in how teachers praise girls and boys in math can reinforce math-gender stereotypes (Corpus & Lepper, 2007; Gunderson et al., 2012). A teacher praising girls exaggeratedly for simple tasks while not praising boys the same way might suggest that the teacher ascribes girls' lower (inherent) abilities than boys. While taking the boys' good performance for granted – likely attributing it to the boys' talent (Sáinz et al., 2020; Tiedemann, 2002) – they reinforce the girls' presumed high(er) effort with strong praise. Through the lens of self-fulfilling prophecies, female learners, due to such potential differences in praise, may adopt the perspective of having weaker math abilities and consider themselves less likely to succeed in math (Jamil et al., 2018). Hence, exaggerated praise for female learners' math performance might reduce female learners' success expectancy and, therefore, motivation (Eccles & Wigfield, 2024).

Research on explicit math-gender stereotype-reinforcing statements is even more scarce. However, such explicit statements might still happen (Leaper & Brown, 2008; Kurtz-Costes et al., 2014). Furthermore, especially for such explicit statements, it would be crucial for teachers to be able to detect and critically evaluate them. Altogether, it would be helpful for teachers to be aware of potential math-gender stereotype-reinforcing behaviors. Such awareness might prevent teachers from behaving that way themselves. Furthermore, it may help teachers to detect and critically evaluate such math-gender stereotype-reinforcing behaviors in students or former teachers.

It, therefore, would be helpful to measure teachers' awareness of math-gender stereotype-reinforcing behaviors. Previous studies on teachers' gender awareness in mathematics and science teaching have applied case vignettes for this purpose (e.g., Atanasova et al., 2024). However, unlike a realistic classroom situation, case vignettes *explicitly* describe *implicit* behaviors. For instance, a difference in tone of voice has to be described explicitly in the vignette; otherwise, it cannot be included. Furthermore, case vignettes can be re-read several times, which allows for the detection of aspects of the vignette that would have otherwise gone unnoticed. These two aspects affect case vignettes' ecological validity. Therefore, teachers' awareness of the abovementioned math-gender stereotype-reinforcing behaviors could be better researched using more ecologically valid measures, for example, *classroom videos*.

## **Studying Teachers' Awareness of Math-gender Stereotype-Reinforcing Behaviors**

### ***The Use of Classroom Videos***

Classroom videos can bridge the theory-practice gap by allowing for the application of professional knowledge in authentic settings (Augsdörfer & Casper, 2018; Augsut et al., 2018; Seidel et al., 2013). As artifacts from practice, they can stimulate teachers' reflections about their own teaching practice (Grossman, 2011). Additionally, classroom videos can assess teachers' noticing, awareness, or informed reasoning (Bergin et al., 2017; Chan et al., 2021; Kersting et al., 2012). Teacher noticing, for instance, is defined as the combination of (1)

selected attention and (2) interpretation in the framework of one's professional knowledge (Chan et al., 2021). This definition is in line with the finding that teachers' awareness of certain teacher behaviors and teaching strategies in classroom videos is related to their teaching knowledge when asking viewers to (1) *detect* and (2) *critically evaluate* suboptimal teacher behaviors and teaching strategies (Kersting et al., 2012). Although both noticing and awareness include detection and interpretation, teacher noticing focuses on student behavior, and teacher awareness focuses on *teacher* behavior.

In classroom video research, both recorded probes of real lessons (Chan et al., 2021) and, more rarely, scripted classroom videos are used to diagnose teacher noticing or awareness (Gartmeier, 2014). Advantages of the latter approach are: First, scripted classroom videos allow showing behaviors that are difficult to record (Gartmeier, 2014), in the case of teachers' math-gender stereotype-reinforcing behaviors possibly due to high self-awareness when being filmed in combination with low social desirability of such behaviors. Second, scripted videos allow for the depiction of suboptimal behaviors and teaching strategies without exposing the depicted individuals. Such depictions of suboptimal behaviors, like math-gender stereotype-reinforcing behaviors, can be of educational value by asking viewers for better behaviors or teaching strategies (Gartmeier, 2014). Third, scripted videos allow for standardization in experimental settings, using the same video scripts and experimentally varying other aspects of the video, like the model teacher and their gender. Hence, in our study, we use scripted classroom videos to use these advantages. Since previous classroom video research has not researched the effects of *gender* on noticing, awareness, or informed reasoning, we study the influence of model teacher gender, participant gender, and model-observer-gender-similarity.

### ***Model Teacher Gender and Participant Gender in Classroom Videos***

Because people socialized as female are more likely to have experienced more discrimination based on the math-gender stereotype than those socialized as male (e.g., Becker & Hall, 2024; Grossman & Porche, 2014; Wang & Degol, 2017), awareness of behaviors that reinforce the math-gender stereotype might be higher among women than among men. Therefore, due to their own experience, women may know that teachers might reinforce the math-gender stereotype and how they might do so through their behaviors. Men, on the contrary, might have difficulties detecting and critically evaluating math-gender stereotype-reinforcing behaviors, as men might (have) benefit(ed) from the math-gender stereotype. Accordingly, people on the more privileged side of a stereotype and hence experience its positive consequences, such as males in math, experience difficulties naming and confronting it (Scheepers et al., 2009). Accordingly, men are less aware of general gender stereotypes (Becker & Swim, 2011). Altogether, different experiences of the genders in educational or math-related contexts may influence participants' awareness of math-gender stereotype-reinforcing behaviors based on their gender.

Additionally, *knowledge* of the different experiences between the genders might influence how participants *perceive* a male or a female model teacher since prior knowledge influences awareness (Chan et al., 2021; Kersting et al., 2012; Kim et al., 2020). Based on their potential prior knowledge of the math-gender stereotype or math-gender stereotype-related discrimination, participants may perceive men as more likely to hold the math-gender stereotype than women. Hence, most people might expect a male (model) teacher but not a female (model) teacher to behave in a way that reinforces the math-gender stereotype. Especially female math teachers, who have likely overcome gender stereotypical barriers to become math teachers themselves (e.g., Avitzour et al., 2020; Eccles, 2011; Gunderson et al., 2012; Rieggle-Crumb & Humpries, 2012; Robinson-Cimpian et al., 2014; Wang & Degol, 2017), might not be expected to reinforce the math-gender stereotype. Hence, participants might be more aware of a male model teacher's math-gender stereotype-reinforcing behaviors than of a female model teacher behaving this way.

### ***Model-Observer-Gender-Similarity***

As gender is a salient social characteristic (Abrams et al., 1990) and an important factor of model-observer-similarity (Bandura, 1977b), own (participant) gender and a model teacher's gender corresponding may improve awareness of math-gender stereotype-reinforcing behaviors (model-observer-gender-similarity). In previous research on instructional videos, model-observer-gender-similarity did not influence *knowledge gains or comprehension* (Hoogerheide et al., 2018). However, as gender is the focus of our current research, and as we confront participants with behaviors that might reinforce the math-gender stereotype, model-observer-gender-similarity will likely matter more than in previous classroom video research.

### **Research Questions, Study Overview and Hypotheses**

To our best knowledge, classroom videos have not been used to assess teachers' awareness of math-gender stereotype-reinforcing behaviors, which leaves the question of whether they provide additional information compared to self-reported questionnaires. They might be beneficial to research the application of (in-)correct knowledge (Seidel et al., 2013) and might be less prone to social desirability than self-reports (van de Mortel, 2008). Further, they may be more ecologically valid than case vignettes (e.g., Atanasova et al., 2024). Hence, this study aims to research teachers' awareness of math-gender stereotype-reinforcing behaviors using classroom videos.

While the role of own gender has been researched in the context of *awareness* of gender stereotypes (Becker & Swim, 2011), such awareness has not been researched for teachers as important socializers in the context of the *math-gender* stereotype that occurs in educational settings (Becker & Hall, 2024; Grossman & Porche, 2014). In our study, we use scripted classroom videos portraying a model teacher's math-gender

stereotype-reinforcing behaviors and vary the model teacher's gender. We further aim to research the effects of model teacher gender, participant gender, and their interaction (model-observer-gender-similarity) on awareness of behaviors that reinforce the math-gender stereotype.

To find out which mechanisms and characteristics, apart from the influence of gender, potentially lie behind awareness of math-gender stereotype-reinforcing behaviors, we further aim to study the relation of awareness of such behaviors with teachers' math-gender stereotype (the global perception of math as male-typed) and teachers' math-gender misconceptions (concrete, incorrect theories justifying girls' inherent lower math abilities).

The hypotheses have been preregistered via [https://aspredicted.org/18R\\_7FY](https://aspredicted.org/18R_7FY). We changed the order of the preregistered hypotheses for this manuscript, reported the results of the two methodological hypotheses (researching the structure and scales of the questionnaire we used) in the methods section (see Study Instruments. The Math-Gender Misconception Questionnaire Short Version), and made minimal changes regarding the wording of the hypotheses. We did not adapt the content of our hypotheses nor change the preregistered analyses after preregistering.

### **Preregistered Gender Hypotheses**

#### ***Model Teacher Gender Hypothesis***

We expect participants to be more aware of the male model teacher's math-gender stereotype-reinforcing behaviors than of the same behaviors in a female model.

#### ***Observing Teacher Gender Hypothesis***

We expect female participants to be more aware of math-gender stereotype-reinforcing behaviors than male participants.

#### ***Model-observer-gender-similarity Hypothesis***

We expect the participating teachers' genders and the model teacher's gender to interact regarding awareness of math-gender stereotype-reinforcing behaviors. If a participant's gender is similar to the model teacher's, we assume that this will improve the participant's awareness.

### **Stereotype- and Misconceptions Hypotheses**

#### ***Math-gender Misconception Hypothesis***

Math-gender *misconceptions* correlate negatively with teachers' awareness of a model teacher's math-gender stereotype-reinforcing behaviors.

#### ***Math-gender Stereotype Hypothesis***

The math-gender *stereotype* correlates negatively with awareness of a model teacher's math-gender stereotype-reinforcing behaviors in math classroom videos.

## Methods

The Local Ethics Committee approved this experiment. We have preregistered the hypotheses and the research plan via [https://aspredicted.org/18R\\_7FY](https://aspredicted.org/18R_7FY). All research data, including syntax for statistical analyses, is publicly available via [https://osf.io/eas4c/?view\\_only=e5bf4721ce8b429fb8055d4ff6518bf1](https://osf.io/eas4c/?view_only=e5bf4721ce8b429fb8055d4ff6518bf1) (converted into a doi upon publication of this manuscript).

### Participants and Recruiting

A total of 278 teachers (169 in-service teachers, 109 pre-service teachers) completed our online experiment without missing data (54 identified as cisgender male, 222 identified as cisgender female, two were nonbinary, but indicated that they were socialized as female;  $M_{age} = 37.01$  [ $SD = 12.23$ , range = 18 – 67 years]). The teachers' school subjects were mostly German ( $n = 151$ ) and math ( $n = 102$ ), followed by other common subjects (English, biology, chemistry and physics, politics and economy, philosophy, geography, other languages, sports, art, or music). About half of the participants (138; 49.6%) studied or taught at least one STEM subject. Many participants studied to teach or already taught at secondary schools ( $n = 166$ ). Participants from all over Germany, Austria, and Switzerland took part in this study.

The communicated topic of the study was "Classroom Video Analysis". Participation was possible between May and July of 2022. We recruited participants via the German-speaking online newspaper News4Teachers, teacher education lectures in different German universities and seminaries, as well as via acquaintances. In total, 578 participants gave informed consent and started the study. Of these, 256 dropped out immediately, on the study's first page. Only 44 participants dropped out during the study, which yielded a final sample of 278 participants. The experimental condition did not affect dropout ( $p > .05$ ).

### Design

This study used a 2x2-(quasi-)experimental design. Participants were randomly assigned to one of two conditions for the experimental factor. Either participant watched four classroom videos of a female model teacher ( $n = 148$ ) or a male model teacher ( $n = 130$ ) expressing math-gender stereotype-reinforcing behaviors. Video scripts and the order of the videos were the same for both conditions (for the Video Scripts, see Appendix A). The second (quasi-experimental) factor was the participating teachers' own gender. We researched the influence of both factors on awareness of math-gender stereotype-reinforcing behaviors.

Furthermore, participants completed a questionnaire on math-gender misconceptions (see Table 1 for the misconception items; see Appendix B for the filler items). We also measured participants' math-gender stereotype using two different measures.

### Study Instruments

### *The Video Task*

Before watching the videos, participants read instructions to watch four short classroom videos of a math lesson and analyze the model teacher's behaviors and their teaching strategies in each video. Participants did not know that math-gender stereotype-reinforcing behaviors would be shown. Like in these short videos, real-life behaviors require teachers' awareness to detect and critically evaluate them without prior warnings or hints. We filmed the videos in a 7<sup>th</sup>-grade classroom of a German secondary school with real teachers and real students who acted according to the video scripts (for scripts, see Appendix A). After filming the videos, we debriefed the students and their teachers. We then conducted a workshop about the math-gender stereotype with the students, discussing their own experiences with teachers' math-gender stereotype-reinforcing behaviors and math-gender stereotyped learning materials, and presenting data on gender similarities in mathematics. We furthermore compensated the class with 150 €, contributing to their next class trip.

Each classroom video was between 0:12 and 2:10 minutes long and presented on a separate page. Each classroom video showed a model teacher's math-gender stereotype-reinforcing behaviors. All participants watched the videos in the same order because the math-gender stereotype-reinforcing behaviors increased in explicitness from videos 1 and 2 (implicit math-gender stereotype-reinforcing behaviors) to videos 3 and 4 (explicit math-gender stereotype-reinforcing behaviors). Specifically, in the first video, the model teacher selected and made up their own math-gender stereotype-reinforcing learning materials (see the example from the introduction). In the second video, the model teacher systematically praised girls exaggeratedly for easy math tasks and praised boys less but appropriately for more complex math tasks. In the third classroom video, the teacher addressed only the boys when motivating their students for a comparative assessment between classes. In the fourth video, the model teacher stated that comparing girls and boys in math was unfair. We based these videos on prior empiric findings and theoretical assumptions about math-gender stereotype-reinforcing behaviors (e.g., Becker & Hall, 2024; Corpus & Lepper, 2007; Eccles, 2011; Gunderson et al., 2021; Jamil et al., 2018; Jehle et al., 2024; Kerkhoven et al., 2016; Moser & Hannover, 2015; Nandi et al., 2024). Each video contained distractors; for example, the model teacher used a random number generator to pick students from the class and answer their questions. Participants could watch each video only once, but they could pause it as often as they wanted. Like this, we aimed at providing ecological validity since real-life behavior can also only be observed once while avoiding participants experiencing time constraints while commenting on the videos and the behaviors they observed. Participants could already answer the open question during video presentation: "List any relevant teaching strategies and teacher behaviors you notice. Which of these behaviors and strategies would you adopt and which not? Give a brief explanation."

First, we measured teachers' *detection* of math-gender stereotype-reinforcing behaviors (1 = detection; 0 = non-detection). Detection was coded if teachers named the math-gender stereotype-reinforcing behavior explicitly as such or if the teacher's answers described the math-gender stereotype-reinforcing behaviors. Second, we coded participants' *critical evaluation* of math-gender stereotype-reinforcing behaviors (-3 to +3). Participants scored a minimum of -3 if they evaluated the math-gender stereotype-reinforcing behaviors *positively* for strong (math-)gender stereotypical reasons, providing more than one argument. Participants scored -2 if they evaluated the math-gender stereotype-reinforcing behaviors positively for weaker (math-)gender stereotypical reasons providing one argument. Participants scored -1 if they evaluated the math-gender stereotype-reinforcing behaviors positively, unrelated to the math-gender stereotype, or without reason. Participants scored a zero if they did not critically evaluate the math-gender stereotype-reinforcing behaviors. Participants scored a +1 if they evaluated the math-gender stereotype-reinforcing behaviors negatively unrelated to the math-gender stereotype or without reason. Participants scored +2 if they evaluated the math-gender stereotype-reinforcing behaviors negatively, arguing with the math-gender stereotype. Participants scored a maximum of +3 if they critically evaluated the math-gender stereotype-reinforcing behaviors, arguing with the math-gender stereotype and discussing negative consequences for female students. Half points were possible (for example answers with coding, see Appendix C).

We summed detection (range: 0 to 1) and critical evaluation (range: -3 to +4) up to *awareness*, as they were neither entirely dependent nor fully independent. For instance, even without detecting a math-gender stereotype-reinforcing behavior, participants may have evaluated it positively (e.g., by suggesting that the selection of [the math-gender stereotype-reinforcing] tasks in the video was positive because the tasks' content was close to students' day-to-day lives). Such an answer would have received a 0 in *detection* and -1 in *critical evaluation*. See Appendix C, for example answers, their scoring and reason for each scoring. See Table 2 for descriptive statistics for detection, critical evaluation, and awareness regarding each video.

Two independent raters quantified participants' open answers. Both raters had a degree in psychology and substantial knowledge about the math-gender stereotype. Both raters were blinded for conditions. The first rater quantified 100 % of all answers regarding the detection and critical evaluation of math-gender stereotype-reinforcing behaviors, and the second rater quantified about 30 % of the answers to assess interrater reliability. Interrater reliability of awareness was good,  $\kappa_{\text{Cohen}} = .77$ ,  $z = 29$ ,  $p < .001$  (Altmann, 1990).

As the participants were asked to mention and explain every behavior and teaching strategy in the model teachers, their answers included information irrelevant to the topic of the math-gender stereotype and thus not analyzed here.

**Follow-up Questions on the Video Analysis.** We included follow-up questions after the video analysis to (a) check if randomization worked and (b) account for differences in the model teachers that we could not control to be similar, such as the model teachers' age. We could also identify covariates to include in our analyses to reduce error variance by statistically controlling for differences (see also Preliminary Analyses).

First, participants reported their active and passive mental effort (Klepsch & Seufert, 2021) because active mental effort can vary in onscreen activities like this experiment (e.g., Rieh et al., 2012; Sidi et al., 2016). Then, participants answered to which degree the model teacher's characteristics, *ethnicity*, *age*, *attitude*, and *gender* attracted their attention on a six-tier Likert scale from *not at all* to *very much*. To avoid positioning effects, these questions on the model teacher's characteristics were presented in a randomized order for each participant. For all follow-up questions, see Appendix D.

### ***The Math-Gender Misconception Questionnaire Short Version***

We assessed teachers' misconceptions with a short version of the Math-Gender Misconceptions Questionnaire (MGMQ-s; see Table 1 for the misconception items). For the new, short version of the MGMQ, we used the two scales most strongly correlated to the math-gender stereotype and measured the two most prevalent math-gender misconceptions (*empathizing-systemizing* and *girls' compensation*; Dersch et al., 2022a). We shortened the 5-item scales by one item each due to substantial content overlap and used the four items with the most potent discriminatory power. Hence, the MGMQ-s consisted of 16 items, eight misconception items (4 per scale), and eight filler items (see Appendix B). An example item for the *empathizing-systemizing* misconception was "As girls think rather empathically and boys think rather systematically, boys are on average more talented in mathematics than girls." An example item for the *girls' compensation* misconception was "For the same grades in mathematics, boys, on average, work less hard than girls, as boys have more natural talent." As all misconception items described scientifically incorrect statements, the correct answer was to disagree with them. Filler items described *true* statements about the math-gender gap, such as "Amongst girls, math is more disliked than amongst boys." For the filler items, the correct answer was to agree with them. They served to balance the questionnaire and to minimize response biases (Moosbrugger & Kelava, 2012; cf. Eitel et al., 2021). Filler items were not analyzed here. Items were presented in a newly randomized order for each participant to control for positioning effects.

Like in the MGMQ, participants first indicated for each item if the statement was *correct* or *incorrect* on a dichotomous scale (i.e., agreement). Then, participants indicated their *certainty* regarding the (in-)correctness of said statement on a five-point Likert scale from *absolutely uncertain* to *absolutely certain* (i.e., certainty). We then calculated misconception scores with the eight misconception items by multiplying agreement (coded with +1) /

disagreement (coded with -1) and response certainty (coded from 0 = *very uncertain* to 4 = *very certain*). Stronger endorsement of a misconception reflects in larger misconception scores (see Table 1 for descriptive values; c.f., Dersch et al., 2022a).

The range of possible values per item was -4 to +4, making it possible to approximate the level of interval-scaled data required to perform confirmatory factor analyses with (robust) maximum likelihood estimation (Eitel et al., 2024). In the preregistered hypotheses about the validation of the questionnaire, we had expected the MGMQ-s' data to fit a two-factor model of math-gender misconceptions (*empathizing-systemizing*, *girls' compensation*) which was supported by the analyses,  $CFI = .977$ ,  $RMSEA = .060$ ,  $SRMR = .039$ ,  $\chi^2 = 43.46$ ,  $df = 19$ ,  $p = .001$ ,  $\chi^2/df = 2.29$ . On the level of local model fit, factor loadings were all significant (all  $ps < .01$ ) and ranged between .40 and .80.

Scale reliabilities were good for *empathizing-systemizing* ( $\omega = .88$ ) and acceptable for *girls' compensation* ( $\omega = .70$ ); for both scales combined, reliability was good ( $\omega = .89$ ).<sup>2</sup>

### ***Math-Gender Stereotype Measures***

Participants first indicated how female-to-male they perceived math from 1 (*very female*) via 5 (*neutral*) to 9 (*very male*) (Dersch et al., 2022a; Nosek et al., 2010; Steffens & Jelenec, 2011). The second measure was giftedness-difference (Steffens & Jelenec, 2011). Participants rated the giftedness of boys/girls for math ("Boys (girls) are often talented for doing math") on a five-tier Likert-scale from *do not agree at all* to *completely agree*. We subtracted the girls' giftedness rating in math from the boys' giftedness rating, with a positive higher value indicating a stronger math-gender stereotype in favor of boys (possible range of -4 to 4; see Steffens & Jelenec, 2011). The two measures were related with a medium effect ( $r = .37$ ,  $p < .001$ ).

### ***Demographics and Other Measures***

We asked for participants' demographics such as age, sex assigned at birth, gender they were socialized as, gender they identify with, native language, teaching (or study) subjects, school type they will work at or already

---

<sup>2</sup> We used R (R Core Team, 2017; version 3.6.23) with the *psych* package (version 2.4.4) for reliability analyses (Revelle, 2017; Revelle & Condon, 2019) and the *lavaan* package (version 0.6-17) for confirmatory factor analyses (Rosseel, 2012). We used maximum likelihood estimation with robust standard errors (MLR) to handle our interval data with moderate deviations from the normal distribution (Li, 2016). For global model fit, we applied the following criteria: a CFI (comparative fit index) value equal to or higher than .95, a root mean square error of approximation (RMSEA) smaller than .06 (Hu & Bentler, 1999), standardized root mean square residual (SRMR) smaller than or equal to .07 (Yu, 2002). Instead of the  $\chi^2$ -test, which describes the discrepancy between the data and the model (Hu & Bentler, 1999), rewarding complex models (West et al., 2012), we rely on the  $\chi^2/df$ -ratio being smaller than 5.0, since this measure penalizes model complexity (Wheaton et al., 1977). Regarding local model fit, the fully standardized factor loadings should be statistically significant ( $p < .05$ ) and higher than .30 (Nunnally, 1978). For robust reliability estimation, we calculated McDonald's omega ( $\omega$ ) (Deng & Chan, 2017). McDonald's omega is acceptable at around 0.70 (Nunnally, 1978), higher than .30 (Nunnally, 1978). For robust reliability estimation, we calculated McDonald's omega ( $\omega$ ) (Deng & Chan, 2017). McDonald's omega is acceptable at around 0.70 (Nunnally, 1978).

work at, and semesters studied as well as their practical experience in teaching. Additionally, we asked for prior occupation with the math-gender stereotype and gender stereotypes in general.

### **Procedure**

After clicking the survey link, participants gave informed consent and were randomly assigned to one of two experimental conditions. Participants watched four videos – one video at a time – of a (male or female) math teacher model's math-gender stereotype-reinforcing behaviors in different classroom situations. They were asked to analyze the teacher's behaviors and teaching strategies. The videos were always in the same order, starting with the video depicting the most implicit math-gender stereotype-reinforcing behaviors and ending with the video depicting the most explicit math-gender stereotype-reinforcing behaviors. Then, participants answered the follow-up questions about the videos. After that, they reported their math-gender misconceptions and the math-gender stereotypes. Finally, participants filled in their demographics and indicated that they had already engaged with the topic of (math-)gender stereotypes. After participation, we thanked the participants and provided a full debriefing. Participants took, on average, 28:12 minutes (Median).

### **Analyses**

We used IBM SPSS statistics, version 26, for data preprocessing and analyzing our main hypotheses (Gender Hypotheses and Misconceptions- and Stereotype Hypotheses). For those hypotheses, we conducted ANCOVAs and Pearson correlations.

## **Results**

### **Teachers' Math-Gender Stereotypes, Math-Gender Misconceptions and Awareness of Math-gender Stereotype-reinforcing Behaviors**

#### ***Teachers' Math-gender Stereotype***

On average, participants rated math to be rather male (9) than female (1) ( $M = 5.41$ ,  $SD = 0.75$ ). A one-sample t-test against the scale-mean of 5 = *neutral*, revealed a medium math-gender stereotype in favor of boys on this measure,  $t(277) = 9.00$ ,  $p < .001$ ,  $d_{\text{Cohen}} = 0.54$  (medium effect; Cohen, 1988). Participants on average indicated slightly higher giftedness-ratings for boys than for girls, leading to a positive, but low giftedness-difference value ( $M = 0.10$ ,  $SD = 0.79$ ; see also Steffens & Jelenec, 2011). A one-sample t-test against 0 = *no talent difference*, revealed a small but significant math-gender stereotype in favor of boys on this measure,  $t(277) = 2.05$ ,  $p = .041$ ,  $d_{\text{Cohen}} = 0.12$  (Cohen, 1988).

#### ***Teachers' Math-gender Misconceptions***

On average, participants did not hold strong math-gender misconceptions according to the general misconception score ( $M = -1.65$ ,  $SD = 1.60$ ). However, agreeing to one incorrect misconception-item per scale

with a relatively high certainty (*certain* or *very certain*) already implies that participants hold a mixed (mis-)conception (e.g., “While Julia is really talented for math, since she is the best of the course, the other girls still lack innate talent because of their gender.”). Mixed (mis-)conceptions still require conceptual change (Vosniadou, 1994; Dersch et al., 2022a; 2022b). We found that 35.3 % of participants had (at least) mixed (mis-)conceptions on at least one of the two misconceptions. As expected and similar to previous research (Dersch et al., 2022a), we found that math-gender misconceptions correlated with both math-gender stereotype measures with a medium effect: (1) the math-female-to-male rating:  $r = .33, p < .001$ , (2) the math-talent-difference measure:  $r = .40, p < .001$ .

### ***Teachers’ Awareness of Math-gender Stereotype-Reinforcing Behaviors***

The first two videos contained implicit math-gender stereotype-reinforcing behaviors: Only 1.8 % of all participants detected the math-gender stereotype-reinforcing behaviors in the first video, and 4.7 % detected it in the second video. The explicit math-gender stereotype-reinforcing behaviors in videos 3 and 4 were detected by 64.4 % and 93.9 % of all participants (see also Table 2). Accordingly, we found low awareness of implicit math-gender stereotype-reinforcing behaviors (first two videos;  $M = 0.02, SD = 0.64$ , on a possible scale of -3 to +4) and medium awareness of explicit math-gender stereotype-reinforcing behaviors (last two videos,  $M = 2.23, SD = 1.02$ , on a possible scale of -3 to +4).

## **Gender Hypotheses**

### **Preliminary Analyses**

The dependent variable of awareness of math-gender stereotype-reinforcing behaviors was approximately normally distributed.

Participants’ gender, STEM profiles and school types as well as the status of being pre-service teacher versus in-service teacher, and German as native language were all equally distributed among the female and male model teacher conditions ( $ps > .05$ ). The conditions did not differ in participants’ age nor teaching experience ( $ps > .05$ ). Furthermore, the conditions did neither differ in their prior occupation with the math-gender stereotype nor with gender stereotypes in general ( $ps > .05$ ). These preliminary results imply that random condition assignment was successful. Technical difficulties were also reported to an equal extent in both groups ( $p > .05$ ). For descriptive values, see Table 3.

Participants’ active mental effort did not differ between the groups,  $t(276) = -0.42, p = .674$ , while it correlated with the dependent variable ( $r = .15, p = .014$ ). This suggests that the intervention did not trigger differences in invested mental effort, but that differences in mental effort influenced awareness as the dependent variable. A large variety in mental effort is likely in online- or on-screen activities, such as this online experiment

(Rieh et al., 2012; Sidi et al., 2016). Therefore, we included mental effort as a covariate to reduce error variance (Field, 2024; Taylor & Innocenti, 1993). We also compared participants' perceptions of the model teacher's age, ethnicity, and attitude between the two conditions. Apart from the model teacher gender, which we had manipulated, we found only the model teachers' perceived attitude to differ significantly between conditions,  $t(276) = -2.84, p = .005$ , and to correlate with awareness of math-gender stereotype-reinforcing behaviors ( $r = .13, p = .030$ ). All the other characteristics did neither differ between groups (all  $ps > .05$ ), nor correlate with the independent variables, meaning model teacher gender and participant gender (all  $ps > .05$ ). Hence, we included the model teacher's attitude as a second covariate to statistically control for perceived differences in attitude between the two model teachers.

### ***Model Teacher Gender Hypothesis***

As expected, we found that participants were more aware of math-gender stereotype-reinforcing behaviors if expressed by a male ( $M = 1.31, SD = 0.65$ ) than by a female model teacher ( $M = 0.96, SD = 0.60$ ),  $F(1, 272) = 23.06, p < .001, \eta^2 = .08$ , with a medium effect.

### ***Observing Teacher Gender Hypothesis***

Contrary to our expectations, female participants ( $M = 1.13, SD = 0.65$ ) did not show a higher awareness of math-gender stereotype-reinforcing behaviors than male participants ( $M = 1.13, SD = 0.65$ ),  $F(1, 272) = 0.31, p = .581$ .

### ***Model-observer-gender-similarity Hypothesis***

Additionally, we expected that if a participant's gender was similar to the model teacher's gender, the participant would be more aware of math-gender stereotype-reinforcing behaviors. Accordingly, we found a small, significant hybrid interaction effect,  $F(1, 272) = 4.97, p = .027, \eta^2 = .02$ . For the interaction pattern, see Figure 1. Simple slope analyses revealed that men's awareness was higher for a male model teacher's math-gender stereotype-reinforcing behaviors than for a female model teacher's math-gender stereotype-reinforcing behaviors ( $p < .001, d_{\text{Cohen}} = 1.03$ ; large effect). While female participants were also more aware of a male model teacher's math-gender stereotype-reinforcing behaviors than of a female model teacher's math-gender stereotype-reinforcing behaviors, the awareness difference between male and female model teacher was small for female participants ( $p = .003, d_{\text{Cohen}} = 0.38$ ). These findings suggest model-observer-gender-similarity, while all participants were more aware of math-gender stereotype-reinforcing behaviors in a male model teacher as expected in the *model teacher gender hypothesis*.

## **Misconceptions and Stereotype Hypotheses**

### ***Math-gender Misconception Hypothesis***

We expected teachers' math-gender misconceptions to be negatively related to their awareness of math-gender stereotype-reinforcing behaviors. We found a small but significant negative correlation of math-gender misconceptions with awareness ( $r = -.16, p = .007$ ).

### ***Math-gender Stereotype Hypothesis***

Regarding the relation of math-gender stereotypes and teachers' awareness of math-gender stereotype-reinforcing behaviors, we found neither significant correlations for the math-female-to-male rating ( $r = .01, p = .883$ ) nor for the talent-difference measure ( $r = -.08, p = .185$ ).

## **Discussion**

This study used scripted classroom videos of model teachers expressing math-gender stereotype-reinforcing behaviors (e.g., selecting stereotypical textbook tasks) to assess participating teachers' awareness of such behaviors. We varied the model teacher's gender (male vs. female). We researched the effect of model teacher gender, participant gender, and the interaction of both (model-observer-gender-similarity) on awareness of math-gender stereotype-reinforcing behaviors. We also researched the relation of awareness of math-gender stereotype-reinforcing behaviors with the math-gender stereotype and math-gender misconceptions.

In brief, we found teachers' awareness to be related to model teachers' gender, model-observer-gender-similarity, and misconceptions but unrelated to teachers' gender as such and teachers' global math-gender stereotype. In the following, we discuss the obtained evidence in detail.

### **Gender Hypotheses**

In the *model teacher gender hypothesis*, we researched if participating teachers were more aware of math-gender stereotype-reinforcing behaviors if they watched the classroom videos with a *male* model teacher than if they watched them with a *female* model teacher expressing such behaviors. As expected, we found this effect. One potential explanation is the knowledge or the social perception of men as less aware towards gender stereotypes (Becker & Swim, 2011; Drury & Kaiser, 2014; Scheepers et al., 2009), leading to the expectation of a male model teacher expressing math-gender stereotype-reinforcing behaviors, while not expecting that same behavior from a female model teacher. Furthermore, a female STEM teacher likely has faced stereotypical barriers herself (e.g., Eccles, 2011; Wang & Degol, 2017), and research finds that female STEM teachers as role models reduce the harmful effects of gender stereotypes for female learners (e.g., Stout et al., 2011). Implicit or explicit knowledge of these mechanisms might make it more difficult to be aware of a *female* STEM teacher's math-gender stereotype-reinforcing behaviors (Chan et al., 2021; Kersting et al., 2012; Kim et al., 2020). While this might be a possible explanation in line with prior research, future research should explicitly assess participants' implicit or explicit knowledge about the math-gender stereotype and its related mechanisms.

In the *observing teacher gender hypothesis*, we researched if the participating teachers' gender influenced their awareness of math-gender stereotype-reinforcing behaviors. We expected awareness to be higher among people socialized as female since they have experienced the math-gender stereotype throughout their time in school and possibly beyond (e.g., Becker & Hall, 2024; Grossman & Porche, 2014). Such experiences with the math-gender stereotype might raise awareness due to prior knowledge (see also Kim et al., 2020). However, we did not find an effect of participants' gender on their awareness. The low awareness for implicit math-gender stereotype-reinforcing behaviors among both genders (see Table 2) suggests the implicit math-gender stereotype-reinforcing behaviors did not stand out enough to mention them in comparison to the distractor behaviors we included in the math classroom videos (e.g., the model teacher using a random numbers generator to pick students to answer their question). The participating teachers' focus is supported by a keyword count, which revealed that  $n = 193$  participants mentioned the random numbers generator in the analysis of Video 2. In comparison, only  $n = 13$  participants mentioned the math-gender stereotype-reinforcing behaviors in the analysis of the same video.

Participating teachers, regardless of gender, being accustomed to implicit math-gender stereotypes, due to their common prevalence in educational settings (e.g., Jehle et al., 2024), may influence this finding. This assumption is furthermore supported by women not indicating more prior concern with the specific *math*-gender stereotype than men,  $t(100.37) = -1.58, p = .117$ . At the same time, women indicated more prior concern with *general* gender stereotypes than men,  $t(93.50) = -2.53, p = .013$ . These findings suggest that, while dealing with and thinking about general gender stereotypes more than men, women do not deal with or think about the *math*-gender stereotype more than men. The math-gender stereotype appears mostly in educational settings and contributes to women not pursuing mathematics or other STEM careers (Eccles, 2011; Wang & Degol, 2017). Female students are likely confronted with the math-gender stereotype during their education. As many female students hold the math-gender stereotype themselves (Cvencek et al., 2011; Steffens et al., 2010), they may have internalized it, likely not questioning it any further. When not pursuing a STEM career, women do not necessarily reflect on said stereotype later in their lives. Thus, women might not necessarily think about that math-gender stereotype more than men if they do not take a STEM career. Accordingly, we found that female teachers who studied to teach or already taught at least one STEM subject ( $n = 113$ ), meaning they pursued a STEM career, dealt with or thought about the math-gender stereotype more than female teachers who did not study any STEM subject  $t(215.32) = -5.20, p < .001$ . Women in a STEM career, henceforth, seem to be forced to think about the math-gender stereotype more due to a more frequent confrontation with it.

In conclusion, our findings suggest that female teachers are not necessarily more aware of math-gender stereotype-reinforcing behaviors than male teachers. These findings align with prior findings, revealing that female

and male student teachers hold math-gender misconceptions to a similar extent (Dersch et al., 2022a). However, both studies used convenience sampling, meaning that unaware male teachers might drop out of a study on the math-gender stereotype, leading to their data not being analyzed.

We also found evidence for the *model-observer-gender-similarity hypothesis*. We found that participant teacher gender and model teacher gender corresponded to improved awareness of math-gender stereotype-reinforcing behaviors. To our knowledge, our study was the first one researching model-observer-gender-similarity in videos other than instructional videos (Hoogerheide et al., 2018). Our findings suggest that gender similarity might play a role in detecting and evaluating another person's gender stereotype-reinforcing behaviors: If one's own gender and the person expressing gender stereotype-reinforcing behavior are similar, this might help detect and dismantle such behaviors. These findings align with learning theories, which suggest model-observer-(gender)-similarity to be beneficial for attention towards a model (Bandura, 1977b). However, our findings reveal model-observer-gender-similarity, especially for men, who were, on average, relatively unaware of the behaviors of a female model teacher (subtly) reinforcing the math-gender stereotype. This unawareness of male teachers of the female model teacher's implicit math-gender stereotype-reinforcing behaviors – additionally to model-observer-gender-similarity – might be influenced by an experienced duality: Men who confront sexism, like math-gender stereotype-reinforcing behaviors, side with women (Drury & Kaiser, 2014), who are mostly the target of sexism (Becker et al., 2014). Confronting a woman for her sexism might feel like not siding with her (anymore). Hence, it might have been harder to critically evaluate or 'call out' a woman's math-gender stereotype-reinforcing behaviors for male participants. Further research applying qualitative methods like interviews is necessary to draw clear conclusions about this possible explanation.

The participating teachers' low awareness of implicit math-gender stereotype-reinforcing behaviors (in the first two videos, see Table 2) appears problematic. Nevertheless, this low awareness does not mean all teachers, being unaware of implicit math-gender stereotype-reinforcing behaviors, would behave math-gender stereotype-reinforcing themselves. To what degree they do needs to be subject to further field research. Regardless, awareness of math-gender stereotype-reinforcing behaviors is important for teachers to detect and dismantle detrimental behaviors that reinforce the math-gender stereotype in other teachers (or students). In sum, our findings call for teacher training on awareness of math-gender stereotype-reinforcing behaviors, especially of *implicit* math-gender stereotype-reinforcing behaviors.

### **Stereotype and Misconceptions Hypotheses**

We found evidence for the *math-gender misconception hypothesis*. The fewer math-gender misconceptions teachers held, the higher their awareness of math-gender stereotype-reinforcing behaviors. The

correlations were small. However, these findings are similar to prior findings of small relations between the implicit and the explicit math-gender stereotype (e.g., Nosek, 2007; Nosek et al., 2010). Similar to studies on implicit and explicit measures, we applied two very different measurement methods: (dis-)agreement to explicit misconception statements in a questionnaire vs. a combination of (1) detection and (2) critical evaluation of math-gender stereotype-reinforcing behaviors in classroom videos. Considering the significant difference between measures, our findings are meaningful.

The relatively small correlation between math-gender misconceptions and awareness of math-gender stereotype-reinforcing behaviors might furthermore suggest that correct theoretical knowledge about the math-gender gap, and hence not holding math-gender misconceptions, helps only little to detect and critically evaluate math-gender stereotype-reinforcing behaviors in (everyday) teaching practice. One explanation might be the theory-practice gap. Even if a participant did not hold strong misconceptions and had correct knowledge (e.g., inflated praise for girls reduces their ability to self-concept), they seem not to have been able to automatically apply all this knowledge to a practical situation (i.e., our detection and evaluation tasks). Being used to gender-stereotyped behaviors and learning materials in school (Becker & Hall, 2024; Grossman & Porche, 2014; Kerkhoven et al., 2016; Moser & Hannover, 2014; Nandi et al., 2024) may have increased this effect.

Given the moderate prevalence of math-gender misconceptions and a relatively high mixed (mis-)conception prevalence of 35.5 %, targeting and hopefully reducing math-gender misconceptions is necessary (see also Dersch et al., 2022a). Reducing math-gender misconceptions might be promising since they are a good starting point for learning (Larkin, 2012) and can be targeted with economic and well-researched interventions such as refutation texts (e.g., Tippett, 2010). Further research is needed to determine if and how the reduction of math-gender misconceptions might benefit awareness of math-gender stereotype-reinforcing behaviors.

In the *math-gender stereotype hypothesis*, we researched the relationship between teachers' math-gender *stereotype* and teachers' awareness of math-gender stereotype-reinforcing behaviors. We expected negative correlations indicating that if a participant reports little math-gender stereotypes, this, on average, leads to more awareness of behaviors that reinforce the math-gender stereotype. However, we did not find this correlation. Our findings suggest that the relation between an explicitly reported math-gender stereotype and awareness of math-gender stereotype-reinforcing behaviors is insignificant. This unrelatedness may be due to the math-gender stereotype measure being more prone to impression management based on social desirability than the measure of math-gender misconceptions. Math-gender stereotypes are self-report attitude measures, whereas math-gender misconceptions measure knowledge, explanations, and justifications for attitudes rather than the attitude themselves (see also Dersch et al., 2022a).

### Limitations and Further Research

An important limitation of the study design is that we assessed the math-gender stereotype after presenting videos of math-gender stereotype-reinforcing behaviors in a math lesson setting. This order may have affected how *male* (rather than female) participants perceived mathematics. We presented the video task before the math-gender stereotype measure to avoid priming effects on teachers' awareness of math-gender stereotype-reinforcing behaviors. If participants had known that the to-be-researched topic was related to mathematics and gender, this would not have allowed for an ecologically valid measurement of awareness of implicit math-gender stereotype-reinforcing behaviors because, in a real classroom setting, teachers also must react to such behaviors without prior warning. Since awareness of math-gender stereotype-reinforcing behaviors was the focus of this research, it was most important for us to avoid priming it. Furthermore, in a prior study, which applied the same measure for the math-gender stereotype, the researched pre-service teachers endorsed the math-gender stereotype to a similar extent as in this study (Dersch et al., 2022a). This referenced study did furthermore not perform any intervention nor show any videos similar to the ones in this study. However, future research should consider order and priming effects and randomly vary stereotype assessment and awareness measure (i.e., the video task).

Furthermore, the math-gender stereotype-reinforcing behaviors in this study are not exhaustive of all possible math-gender stereotype-reinforcing behaviors. Therefore, the measured awareness cannot fully represent all potential awareness of math-gender stereotype-reinforcing behaviors. We did, however, portray a broad range of implicit and explicit math-gender stereotype-reinforcing behaviors to provide a first overview of such situations requiring teachers' awareness. Future studies could apply different classroom videos showing math-gender stereotype-reinforcing behaviors to research other aspects of awareness of them.

Another limitation of our study is that we did not measure participants' own math-gender stereotype-reinforcing behaviors but their *awareness* of math-gender stereotype-reinforcing behaviors in scripted math classroom videos, which is not a behavioral measure. However, applying such scripted classroom videos can help bridge the theory-practice gap to some extent (e.g., August et al., 2018).

A limitation regarding this study's sample is that only few participants socialized as male took part in the study ( $n = 54$ ), while most participants were socialized as female ( $n = 224$ ). However, participants socialized as male vs. female were equally distributed among both conditions (see Preliminary Analyses and Table 3). While the gender distribution in our experiment is not representative of society, it is mostly representative of teachers in Germany: According to the Federal Office of Statistics of Germany, in the school year of 2019 / 2020, 73.0 % of teachers in schools for general education were female. In our sample, 80.6 % of participants were female (Destatis, 2003).

Finally, this study did not consider stereotypes based on ethnicity, which are also researched to negatively impact stereotyped students in mathematics (e.g., Copur-Gencturk et al., 2023; Riegle-Crumb & Humpries, 2012). Future research should take the intersectionality of (math-)gender stereotypes overlapping with other forms of discrimination, such as social class, race, or ethnicity, into account and aim at targeting and reducing the harmful effects of such intersectionality (Keller et al., 2023; Yuval-Davis, 2006).

### **Conclusion**

We found that teachers' awareness of implicit math-gender-stereotype-reinforcing behaviors is low. However, awareness of such behaviors is important to provide equal chances for students of all genders. Only teachers aware of such behaviors can modify their own behavior if necessary and inspire others to critically reflect on their potentially stereotype-reinforcing behaviors in school. Interventions improving awareness are thus necessary.

Additionally, this awareness is related to math-gender misconceptions, which contribute to the manifestation of the math-gender gap (see also Dersch et al., 2022a; Girelli, 2023; Rubel et al., 2024). Further research should investigate if findings regarding these relations can be replicated and how interventions targeting math-gender misconceptions might affect awareness of math-gender stereotype-reinforcing behaviors or vice versa.

### References

- Abrams, D, Thomas, J., & Hogg, M. A. (1990). Numerical distinctiveness, social identity and gender salience. *British Journal of Social Psychology*, 29(1), 87-92. <https://doi.org/10.1111/j.2044-8309.1990.tb00889.x>
- Altman, D. G. (1990). *Practical statistics for medical research*. Chapman and Hall/CRC.
- Ashmore, R. D., & Del Boca, F. K. (1979). Sex stereotypes and implicit personality theory: Toward a cognitive—Social psychological conceptualization. *Sex roles*, 5(2), 219-248. <https://doi.org/10.1007/BF00287932>
- Atanasova, S., Robin, N., & Brovelli, D. (2024). Interest, learning opportunities and teaching experience as predictors of professional vision in gender-sensitive physics education. *International Journal of Science Education*, 1-23. <https://doi.org/10.1080/09500693.2024.2406528>
- Augsdörfer, A., & Casper, M. (2018). Closing the Theory-Practice Gap: Employing Authentic Video-Taped Lessons in Vocational Teacher Education. In C. Nägele & B. E. Stalder (Eds), pp.1–8. <https://doi.org/10.5281/zenodo.1319622>
- August, J. N., Boyd, L. D., & Giblin-Scanlon, L. (2018). Bridging the theory-practice gap with dental hygiene instrumentation videos. *Journal of dental education*, 82(9), 961-967. <https://doi.org/10.21815/JDE.018.095>
- Authors (submitted). Thinking is not Telling: Student Teachers’ Implicit Math-gender Association is Unrelated to their Explicit Math-Gender Stereotype.
- Avitzour, E., Choen, A., Joel, D., & Lavy, V. (2020). *On the origins of gender-biased behavior: The role of explicit and implicit stereotypes*. National Bureau of Economic Research. <https://doi.org/10.3386/w27818>
- Bandura, A. (1977a). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- Bandura, A. (1977b). *Social Learning Theory*. Prentice Hall.
- Bandura, A., (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall, Inc.
- Baron-Cohen, S. (2005, January). The essential difference: The male and female brain. *Phi Kappa Phi Forum*, 85(1), 23 – 26.
- Becker, J. R., & Hall, J. (2024). Research on gender and mathematics: Exploring new and future directions. *ZDM—Mathematics Education*, 56(1), 141-151. <https://doi.org/10.1007/s11858-023-01510-6>

- Becker, J. C., & Swim, J. K. (2011). Seeing the unseen: Attention to daily encounters with sexism as way to reduce sexist beliefs. *Psychology of Women Quarterly, 35*(2), 227-242.  
<https://doi.org/10.1177/0361684310397509>
- Becker, J. C., Zawadzki, M. J., & Shields, S. A. (2014). Confronting and reducing sexism: A call for research on intervention. *Journal of Social Issues, 70*(4), 603-614. <https://doi.org/10.1111/josi.12081>
- Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences, 107*(5), 1860-1863.  
<https://doi.org/10.1073/pnas.0910967107>
- Bergin, C., Wind, S. A., Grajeda, S., & Tsai, C. L. (2017). Teacher evaluation: Are principals' classroom observations accurate at the conclusion of training?. *Studies in Educational Evaluation, 55*, 19-26.  
<https://doi.org/10.1016/j.stueduc.2017.05.002>
- Chan, K. K. H., Xu, L., Cooper, R., Berry, A., & van Driel, J. H. (2021). Teacher noticing in science education: do you see what I see?. *Studies in Science Education, 57*(1), 1-44.  
<https://doi.org/10.1080/03057267.2020.1755803>
- Cimpian, J. R., Lubienski, S. T., Timmer, J. D., Makowski, M. B., & Miller, E. K. (2016). Have gender gaps in math closed? Achievement, teacher perceptions, and learning behaviors across two ECLS-K cohorts. *AERA Open 2*:23328588416673617. <https://doi.org/10.1177/2332858416673617>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum.
- Corpus, J. H., & Lepper, M. R. (2007). The effects of person versus performance praise on children's motivation: Gender and age as moderating factors. *Educational Psychology, 27*, 487-508.  
<https://doi.org/10.1080/01443410601159852>
- Deng, L., & Chan, W. (2017). Testing the difference between reliability coefficients alpha and omega. *Educational and psychological measurement, 77*(2), 185-203.  
<https://doi.org/10.1177/0013164416658325>
- Copur-Gencturk, Y., Thacker, I., & Cimpian, J. R. (2023). Teachers' race and gender biases and the moderating effects of their beliefs and dispositions. *International Journal of STEM Education, 10*(1), 1 - 25.  
<https://doi.org/10.1186/s40594-023-00420-z>
- 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>
- Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2012). Balanced identity theory. *Cognitive consistency: A fundamental principle in social cognition, 157-177*.

- Dersch, A.-S., Heyder, A., & Eitel, A. (2022a). Exploring the Nature of Teachers' Math-gender Stereotypes: The Math-gender Misconception Questionnaire. *Frontiers in psychology*, 1728. <https://doi.org/10.3389/fpsyg.2022.820254>
- Dersch, A.-S., Renkl, A., & Eitel, A. (2022b). Personalized refutation texts best stimulate teachers' conceptual change about multimedia learning. *Journal of Computer Assisted Learning*. <https://doi.org/10.1111/jcal.12671>
- Destatis (2023, July 4). Schulen: Lehrkräfte nach Schulart und Beschäftigungsumfang. *Datenreport 2021* [Schools: Teachers Regarding School Type and Scope of Employment. *Data report 2021*. [https://www.destatis.de/DE/Service/Statistik-Campus/Datenreport/Downloads/datenreport-2021-kap-3.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Service/Statistik-Campus/Datenreport/Downloads/datenreport-2021-kap-3.pdf?__blob=publicationFile)
- Duckworth, A. L., & Seligman, M. E. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of educational psychology*, 98(1), 198. <https://psycnet.apa.org/doi/10.1037/0022-0663.98.1.198>
- Duguid, M. M., & Thomas-Hunt, M. C. (2015). Condoning stereotyping? How awareness of stereotyping prevalence impacts expression of stereotypes. *Journal of Applied Psychology*, 100(2), 343–359. <https://doi.org/10.1037/a0037908>
- Drury, B. J., & Kaiser, C. R. (2014). Allies against sexism: The role of men in confronting sexism. *Journal of social issues*, 70(4), 637-652. <https://doi.org/10.1111/josi.12083>
- Eagly, A. H., & Koenig, A. M. (2021). The vicious cycle linking stereotypes and social roles. *Current Directions in Psychological Science*, 30(4), 343-350. <https://doi.org/10.1177/09637214211013775>
- Eccles, J. (2011). Gendered educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *International Journal of Behavioral Development*, 35(3), 195-201. <https://doi.org/10.1177/0165025411398185>
- Eccles, J., Adler, T. F., Futtermann, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectations, values and academic behaviors in Perspective on achievement and achievement motivation. In J. T. Spence (Ed.), pp. 75–146, Academic Press.
- Eccles, J. S., & Wigfield, A. (2024). The development, testing, and refinement of Eccles, Wigfield, and colleagues' situated expectancy-value model of achievement performance and choice. *Educational Psychology Review*, 36(2), 51. <https://doi.org/10.1007/s10648-024-09888-9>

- Eitel, A., Prinz, A., Kollmer, J., Niessen, L., Russow, J., Ludäscher, M., Renkl, A., & Lindner, M. A. (2021). The misconceptions about multimedia learning questionnaire. An empirical evaluation study with teachers and student teachers. *Psychology Learning and Teaching, 20*(3), 420 – 444.  
<https://doi.org/10.1177%2F14757257211028723>
- Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103–127.  
<https://psycnet.apa.org/doi/10.1037/a0018053>
- Eliot, L. (2010). *Pink brain, blue brain: How small differences grow into troublesome gaps-and what we can do about it*. Simon and Schuster. <https://doi.org/10.14507/er.v0.1310>
- Ellemers, N. (2018). Gender Stereotypes. *Annual Review of Psychology, 69*, 275–298.  
<https://doi.org/10.1146/annurev-psych-122216-011719>
- Escovar, E., Rosenberg-Lee, M. Uddin, L. Q., & Menon, V. (2016). The empathizing-systemizing theory, social abilities, and mathematical achievement in children. *Scientific reports, 6*(1), 1 – 11.  
<https://doi.org/10.1038/srep23011>
- Field, A. (2024). *Discovering statistics using IBM SPSS statistics*. Sage Publications Limited.
- Fiske, S. T. (1982). Schema-triggered affect: Applications to social perception. In *Affect and cognition: 17th Annual Carnegie Mellon symposium on cognition* (pp. 55-78). Hillsdale: Lawrence Erlbaum.
- Gartmeier, M. (2014). Fiktionale Videofälle in der Lehrerinnen-und Lehrerbildung [Fictional Videos for Teacher Education]. *Beiträge zur Lehrerinnen-und Lehrerbildung, 32*(2), 235-246.  
<https://doi.org/10.25656/01:13868>
- Girelli, L. (2023). What does gender has to do with math? Complex questions require complex answers. *Journal of Neuroscience Research*. <https://doi.org/10.1002/jnr.25056>
- Gottfredson, L. S. (1998). The general intelligence factor.
- Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological review, 109*(1), 3. <https://psycnet.apa.org/doi/10.1037/0033-295X.109.1.3>
- Grossman, P. (2011). Framework for teaching practice: A brief history of an idea. *Teachers College Record, 113*(12), 2836-2843. <https://doi.org/10.1177/016146811111301205>
- Grossman, J. M., & Porche, M. V. (2014). Perceived gender and racial/ethnic barriers to STEM success. *Urban Education, 49*(6), 698-727. <https://doi.org/10.1177/0042085913481364>

- Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The Role of Parents and Teachers in the Development of Gender-Related Math Attitudes. *Sex Roles, 66*, 153–166 (2012).  
<https://doi.org/10.1007/s11199-011-9996-2>
- Haines, E. L., Deaux, K., & Lofaro, N. (2016). The times they are a-changing... or are they not? A comparison of gender stereotypes, 1983–2014. *Psychology of Women Quarterly, 40*(3), 353-363.  
<https://doi.org/10.1177/0361684316634081>
- Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychol. Sci. Public Inter. 8*, 1–51.  
<https://10.1111/j.1529-1006.2007.00032.x>
- Hand, S., Rice, L., & Greenlee, E. (2017). Exploring teachers' and students' gender role bias and students' confidence in STEM fields. *Social Psychology of Education, 20*(4), 929–945.  
<https://doi.org/10.1007/s11218-017-9408-8>
- Holder, K., & Kessels, U. (2017). Gender and ethnic stereotypes in student teachers' judgments: A new look from a shifting standards perspective. *Social psychology of education, 20*, 471-490.  
<https://doi.org/10.1007/s11218-017-9384-z>
- Hoogerheide, V., van Wermeskerken, M., van Nassau, H., & van Gog, T. (2018). Model-observer similarity and task-appropriateness in learning from video modeling examples: Do model and student gender affect test performance, self-efficacy, and perceived competence?. *Computers in Human Behavior, 89*, 457-464. <https://doi.org/10.1016/j.chb.2017.11.012>
- Hu, L.T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1-55.  
<https://doi.org/10.1080/10705519909540118>
- Hughes, S., Lyddy, F., & Lambe, S. (2013). Misconceptions about Psychological Science: A Review. *Psychology Learning & Teaching, 12*, 20 - 31. <https://doi.org/10.2304/2Fplat.2013.12.1.20>
- Jamil, F. M., Larsen, R. A., & Hamre, B. K. (2018). Exploring longitudinal changes in teacher expectancy effects on children's mathematics achievement. *Journal for Research in Mathematics Education, 49*(1), 57-90. <https://doi.org/10.5951/jresmetheduc.49.1.0057>
- Jehle, A. M., Groeneveld, M. G., van de Rozenberg, T. M., & Mesman, J. (2024). The hidden lessons in textbooks: Gender representation and stereotypes in European mathematics and language books. *European Journal of Education, 59*(4), e12716. <https://doi.org/10.1111/ejed.12716>

- Jussim, L., Harber, K. D., Crawford, J. T., Cain, T. R., & Cohen, F. (2005). Social reality makes the social mind: Self-fulfilling prophecy, stereotypes, bias, and accuracy. *Interaction Studies*, *6*(1), 85-102.  
<https://doi.org/10.1075/is.6.1.07jus>
- Kahn, S., & Ginther, D. (2017). *Women and STEM*. National Bureau of Economic Research. <https://doi.org/10.3386/w23525>
- Keller, C. (2001). Effect of Teachers' Stereotyping on Students' Stereotyping of Mathematics as a Male Domain. *The Journal of Social Psychology*, *141*(2), 165–173.  
<https://doi.org/10.1080/00224540109600544>
- Keller, L., Luedtke, O., Preckel, F., & Brunner, M. (2023). Educational Inequalities at the Intersection of Multiple Social Categories: An Introduction and Systematic Review of the Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy (MAIHDA) Approach. *Educational Psychology Review*, *35*(1), 31. <https://doi.org/10.1007/s10648-023-09733-5>
- Kerkhoven, A. H., Russo, P., Land-Zandstra, A. M., Saxena, A., & Rodenburg, F. J. (2016). Gender stereotypes in science education resources: A visual content analysis. *PloS one*, *11*(11), e0165037.  
<https://doi.org/10.1371/journal.pone.0165037>
- Kersey, A. J., Csumitta, K. D., & Cantlon, J. F. (2019). Gender similarities in the brain during mathematics development. *npj Science of Learning*, *4*(1), 1 – 7. <https://doi.org/10.1038/s41539-019-0057-x>
- Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, *49*(3), 568-589.  
<https://doi.org/10.3102/0002831212437853>
- Kim, S., & Rehder, B. (2011). How prior knowledge affects selective attention during category learning: An eyetracking study. *Memory & cognition*, *39*, 649-665.
- Klepsch, M. & Seufert, T. (2021). Making an effort versus experiencing load. In *Frontiers in Education*, *6*, 645284. <https://doi.org/10.3389/educ.2021.645284>
- Klineberg, O. (1951). The scientific study of national stereotypes. *International social science bulletin*, *3*(3), 505-514.
- Kollmayer, M., Schultes, M. T., Lüftenegger, M., Finsterwald, M., Spiel, C., & Schober, B. (2020, July). REFLECT—A teacher training program to promote gender equality in schools. In *Frontiers in Education* (Vol. 5, p. 136). Frontiers Media SA. <https://doi.org/10.3389/educ.2020.00136>

- Kurtz-Costes, B., Copping, K. E., Rowley, S. J., & Kinlaw, C. R. (2014). Gender and age differences in awareness and endorsement of gender stereotypes about academic abilities. *European Journal of Psychology of education, 29*, 603-618.  
<https://doi.org/10.1007/s10212-014-0216-7>
- Lachance, J. A., & Mazzocco, M. M. (2006). A longitudinal analysis of sex differences in Math and spatial skills in primary school age children. *Learning and individual differences, 16*(3), 195 – 2016.  
<https://doi.org/10.1016/j.lindif.2014.03.020>
- Larkin, D. (2012). Misconceptions about “misconceptions”: Preservice secondary science teachers’ views on the value and role of student ideas. *Science Education, 96*(5), 927-959. <https://doi.org/10.1002/sce.21022>
- Leaper, C., & Brown, C. S. (2008). Perceived experiences with sexism among adolescent girls. *Child development, 79*(3), 685-704. <https://doi.org/10.1111/j.1467-8624.2008.01151.x>
- Li, Q. (1999). Teachers’ beliefs and gender differences in mathematics: a review. *Educational Research, 41*(1), 63–76. <https://doi.org/10.1080/0013188990410106>
- Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behav. Res. Methods 48, 936–949*. <https://doi.org/10.3758/s13428-015-0619-7>
- Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychological bulletin, 136*(6), 1123.  
<https://psycnet.apa.org/doi/10.1037/a0021276>
- Martingano, A. J., & Konrath, S. (2022). How cognitive and emotional empathy relate to rational thinking: empirical evidence and meta-analysis. *The Journal of Social Psychology, 162*(1), 143–160.  
<https://doi.org/10.1080/00224545.2021.1985415>
- Moosbrugger, H., & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion [Testing Theory and Questionnaire Construction]*. [https://doi.org/10.1007/978-3-642-20072-4\\_1](https://doi.org/10.1007/978-3-642-20072-4_1)
- Moser, F., & Hannover, B. (2014). How gender fair are German schoolbooks in the twenty first century? An analysis of language and illustrations in schoolbooks for mathematics and German. *European journal of psychology of education, 29*(3), 387-407. <https://doi.org/10.1007/s10212-013-0204-3>
- Nandi, A., Halder, T., & Das, T. (2024). Gender Stereotypes and Bias in School Mathematics Textbooks: A Systematic Literature Review. *Current Perspectives in Educational Research, 7*(2), 36-50. <https://doi.org/10.46303/cuper.2024.8>

- Nosek, B. A. (2007). Implicit–explicit relations. *Current directions in psychological science*, *16*(2), 65-69.  
<https://doi.org/10.1111%2Fj.1467-8721.2007.00477.x>
- Nosek, B. A., & Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *American Educational Research Journal*, *48*(5), 1125-1156.  
<https://doi.org/10.3102%2F0002831211410683>
- Nosek, B.A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., & Banaji, M. R. (2010). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*(1), 36-88.  
<https://doi.org/10.1080/10463280701489053>
- Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.).
- 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*, 645–662. <https://doi.org/10.1037/edu0000012>
- Revelle, W. R. (2017). psych: Procedures for Personality and Psychological Research. Software
- Revelle, W., & Condon, D. M. (2019). Reliability from  $\alpha$  to  $\omega$ : a tutorial. *Psychol. Assess.*, *31*, 1395.  
<https://doi.org/10.1037/pas0000754>
- Riegle-Crumb, C., & Humphries, M. (2012). Exploring bias in math teachers' perceptions of students' ability by gender and race/ethnicity. *Gender & Society*, *26*(2), 290-322.  
<https://doi.org/10.1177/0891243211434614>
- Rieh, S. Y., Kim, Y. M., & Markey, K. (2012). Amount of invested mental effort (AIME) in online searching. *Information Processing & Management*, *48*(6), 1136-1150. <https://doi.org/10.1016/j.ipm.2012.05.001>
- Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., & Copur-Gencturk, Y. (2014). Teachers' perceptions of students' mathematics proficiency may exacerbate early gender gaps in achievement. *Developmental psychology*, *50*(4), 1262. <https://doi.org/10.1037/a0035073>
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of statistical software*, *48*(2), 1–36.
- Rubel, L. H., Ayalon, M., & Shahbari, J. A. (2024). Sequencing & selecting solutions in a gendered world. *Mathematical Thinking and Learning*, *26*(2), 185-207. <https://doi.org/10.1080/1098606>

- Sáinz, M., Fàbregues, S., & Solé, J. (2020). Parent and teacher depictions of gender gaps in secondary student appraisals of their academic competences. *Frontiers in Psychology, 11*, 573752.  
<https://doi.org/10.3389/fpsyg.2020.573752>
- Scheepers, D., Ellemers, N., & Sintemaartensdijk, N. (2009). Suffering from the possibility of status loss: Physiological responses to social identity threat in high status groups. *European Journal of Social Psychology, 39*(6), 1075-1092. <https://doi.org/10.1002/ejsp.609>
- Seidel, T., Blomberg, G., & Renkl, A. (2013). Instructional strategies for using video in teacher education. *Teaching and Teacher Education, 34*, 56-65. <https://doi.org/10.1016/j.tate.2013.03.004>
- Sidi, Y., Ophir, Y., & Ackerman, R. (2016). Generalizing screen inferiority-does the medium, screen versus paper, affect performance even with brief tasks?. *Metacognition and Learning, 11*, 15-33.  
<https://doi.org/10.1007/s11409-015-9150-6>
- Starr, C. R., & Simpkins, S. D. (2021). High school students' math and science gender stereotypes: relations with their STEM outcomes and socializers' stereotypes. *Social Psychology of Education, 24*, 273-298.  
<https://doi.org/10.1007/s11218-021-09611-4>
- 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), 324-335. <https://doi.org/10.1007/s11199-010-9924-x>
- Steffens, M. C., Jelenec, P., & Noack, P. (2010). On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *Journal of Educational Psychology, 102*(4), 947-963. <https://doi.org/10.1037/a0019920>
- Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of personality and social psychology, 100*(2), 255.  
<https://psycnet.apa.org/doi/10.1037/a0021385>
- Taylor, M. J., & Innocenti, M. S. (1993). Why covariance? A rationale for using analysis of covariance procedures in randomized studies. *Journal of early Intervention, 17*(4), 455-466.  
<https://doi.org/10.1177/105381519301700409>
- Tiedemann, J. (2002). Teachers' gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educational Studies in mathematics, 50*(1), 49-62.  
<https://doi.org/10.1023/A:1020518104346>

- Tippett, C. D. (2010). Refutation text in science education: A review of two decades of research. *International Journal of Science and Mathematics Education*, 8(6), 951-970. <https://doi.org/10.1007/s10763-010-9203>
- Wang, M. T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review*, 29, 119 – 140. <https://doi.org/10.1007/s10648-015-9355-x>
- Wheaton, B. (1977). Assessing reliability and stability in panel models. *Sociological methodology/Jossy-Bass*. <https://doi.org/10.2307/270754>
- van de Mortel, T. F. (2008). Faking it: social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, The*, 25(4), 40-48. <https://doi.org/10.3316/informit.210155003844269>
- Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and instruction*, 4(1), 45-69. [https://doi.org/10.1016/0959-4752\(94\)90018-3](https://doi.org/10.1016/0959-4752(94)90018-3)
- West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). The Guilford Press.
- Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. University of California, Los Angeles.
- Yuval-Davis, N. (2006). Intersectionality and feminist politics. *European journal of women's studies*, 13(3), 193-209. <https://doi.org/10.1177%2F1350506806065752>

**Table 1**

*Descriptive Values of the Short Version of the Math-gender Misconception Questionnaire (MGMQ-s).*

	Agreement rates (min. = 0, max. = 1)	Response certainty (min. = 0, max. = 4)	Misconception score <sup>a</sup> (min. = -4, max. = +4)	Item-total correlation (min. = 0, max. = 1)
<b>Empathizing-Systemizing (ES):</b> $\omega = .88$				
ES1: <i>Mathematical correlations are usually easier for boys to understand than for girls, because boys tend to think in systematic contexts</i>	.24	2.61 (1.03)	-1.66 (2.26)	.72
ES2: <i>As boys, more likely think in systematic categories, they have better cognitive prerequisites for math than girls do.</i>	.25	2.49 (1.06)	-1.56 (2.22)	.67
ES3: <i>Female empathy makes it easier for girls to deal with people, while boys are on average more gifted in systematic thinking and thus in mathematics.</i>	.30	2.59 (1.00)	-1.26 (2.48)	.74
ES4: <i>On average, girls think more empathically than boys do, boys on the other hand are more talented in systematic thinking and thus also in mathematics.</i>	.35	2.52 (0.96)	-1.07 (2.48)	.70
<b>Girls' Compensation (GC):</b> $\omega = .70$				
GC1: <i>Girls normally have to work harder to perform equally well in mathematics as boys.</i>	.23	2.78 (0.96)	-1.70 (2.40)	.39
GC2: <i>Girls compensate for their on average lower aptitude in mathematics compared to boys by being more diligent.</i>	.35	2.64 (0.99)	-1.00 (2.64)	.46
GC3: <i>Mathematical content often just flies to boys, while girls have to make more effort for it on average.</i>	.15	2.78 (0.89)	-2.14 (2.00)	.54
GC4: <i>To achieve equally good grades in mathematics, boys have to make less effort because they are more talented than girls are.</i>	.07	3.05 (0.92)	-2.81 (1.49)	.47
<b>All Items:</b> $\omega = .89$				

*Note.* Agreement rates represent the proportion of participants agreeing to a statement. Descriptive values for response certainty and misconception scores represent means and standard deviations (in parentheses).

<sup>a</sup>Calculated by converting agreement into +1 and disagreement into -1, then multiplied with response certainty.

**Table 2**

*Descriptive Values of Teachers' Detection, Critical Evaluation of and Awareness Of Math-gender Stereotype-reinforcing Behaviors (MGSRB).*

Video	Detection of MGSRB		Critical Evaluation of MGSRB (score: -3 to +3)		Awareness of MGSRB (score: - 3 to +4)	
	Frequency	%	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Video 1 (implicit)	5	1.8	0.09	0.44	0.11	0.54
Video 2 (implicit)	13	4.7	-0.11	0.99	-0.06	1.10
Video 3 (explicit)	179	64.4	1.21	1.18	1.86	1.59
Video 4 (explicit)	261	93.9	1.66	0.91	2.60	1.04

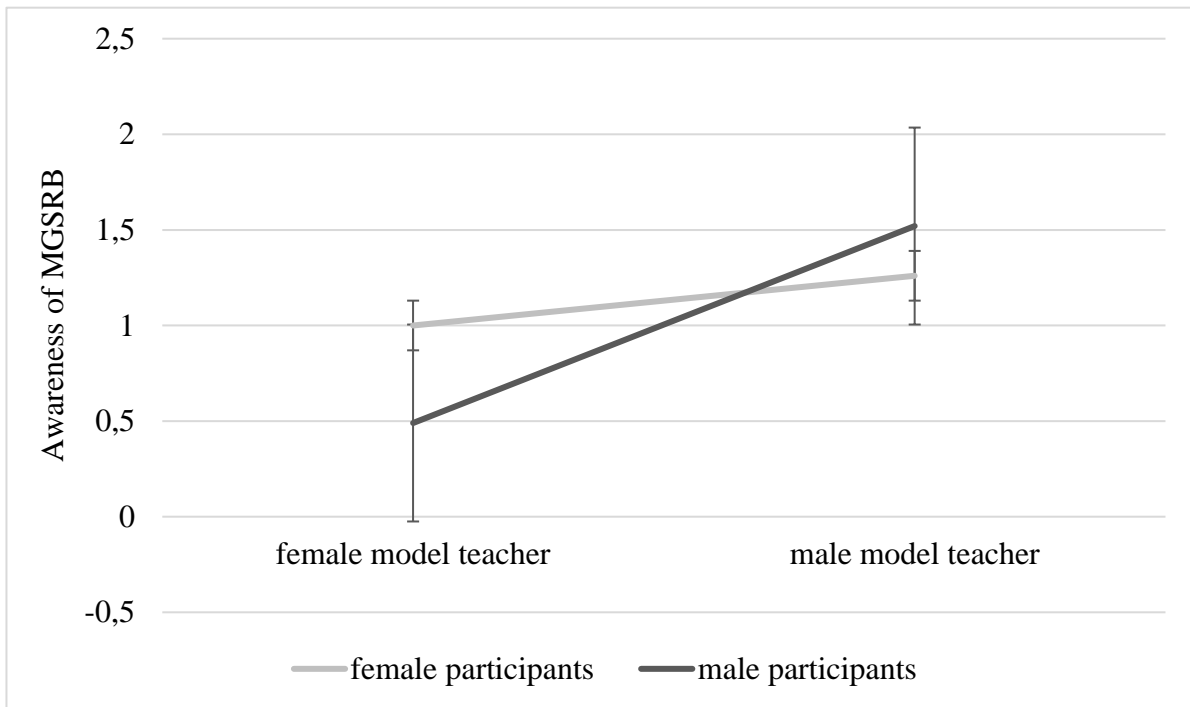
*Note.* Awareness is the sum of Detection (0; 1) and Critical Evaluation (-3 to +4). Hence, the possible score for Awareness is -3 to +4

**Table 3***Participant Characteristics between Conditions*

	Female Model Teacher		Male Model Teacher	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Age	37.60	12.53	36.34	11.90
Teaching experience (1 = none – 4 = examined full-time school teacher)	2.77	0.93	2.85	0.90
	% (a)	% (b)	% (a)	% (b)
Gender socialized as (a) female, (b) male	76.4 %	22.3 %	83.8 %	16.2 %
Status (a) pre-service teacher, (b) in-service teacher	39.9 %	60.1 %	38.5 %	61.5 %
School type (a) primary school, (b) secondary school	37.2 %	62.8 %	43.8 %	56.2 %
STEM profile (a) yes, (b) no	49.3 %	50.7 %	50.0 %	50.0 %
German as native language (a) yes, (b) no	94.6 %	5.4 %	96.2 %	3.8 %
Technical difficulties while watching the video (a) yes, (b) no	12.8 %	88.2 %	7.7 %	92.3 %

**Figure 1**

*Pattern of the Interaction Between Model Teacher Gender and Participant Teacher Gender on Mean Awareness of Math-gender Stereotype-reinforcing Behaviors*



**Appendices****Appendix A**

*Video Scripts (all filmed with female and male model teacher)*

Video 1 (implicit)

*Students are seen from behind. Teacher stands in front of the class with open math textbook and says:* Ah,

here, the task of Mr. Meier, the road engineer, it's quite good. He needs  $\frac{1}{4}$

ton of tar and 300kg of concrete. How many tons of concrete does he need then? The result please as a fraction, please...Yes, Thomas..?

*Thomas:*  $\frac{3}{10}$  t of concrete.

*Teacher:* That is correct. Now I've come up with a similar task for you: Mrs. Schäfer goes to the market and needs another  $\frac{1}{3}$  kg of apples and  $\frac{1}{2}$  kg of flour for her cake recipe. How many grams does she need each?

Teacher turns around and writes values ( $\frac{1}{3}$  and  $\frac{1}{2}$ ) on the board. Teacher: Yes... Lisa?

Video 2 (implicit):

*Students are seen from behind. Teacher stands at the blackboard and sets tasks. They use a generator app on their phone to generate random numbers between 1 and 30.*

*Teacher:* Number 24, (looks at class list....) Lisa, What does  $\frac{3}{4} + \frac{5}{4}$  add up to? Lisa:  $\frac{8}{4}$ . *Teacher:* Wow, you did a great job. Very good.

Number 13... Jonas? How would you express that correctly?

*Jonas:*  $\frac{8}{4}$  are actually two.

*Teacher:* Yes, exactly. I'll draw it for you.

*Teacher draws a circle, with a cross through it to suggest quarters, wipes out third quarter. Teacher:* Here we have  $\frac{3}{4}$ . Now add  $\frac{5}{4}$ .

*Teacher draws fourth quarter again, saying:*

One,

*draws another circle and draws a cross through it counts:*

2, 3, 4, 5.

So Lisa, how many circles do you see here?

*Lisa:* Two.

*Teacher:* Yes, exactly, very good!

Let's take Number 2 next, Steffen; what is the smallest common denominator of  $\frac{20}{48}$  and  $\frac{1}{4}$ ?

*Steffen:*  $\frac{20}{48}$  can be shortened by 4. Then it's  $\frac{5}{12}$ .  $\frac{1}{4}$  can be expanded by 3, then we have  $\frac{3}{12}$ . Accordingly,

12 is the lowest common denominator.

*Teacher:* Correct. Fine, let's move on.

Number 16: Johanna, can you please expand  $\frac{1}{8}$  so that it has the same denominator as  $\frac{1}{32}$ ?

*Johanna:* Yes, I would then expand it with 4. That would make  $\frac{4}{32}$ . *Teacher:* Wow, you did that very well, great, Johanna!

Video 3 (explicit):

*Teacher enters the classroom, students not visible. Teacher stands in front of the class, students are seen from behind. Teacher puts bag on table and begins to speak.*

*Teacher:* Sooo, today we will write an unannounced math test. This will be a comparison test between all of year 7. Boys, *(looks in the direction of a group of boys, past the girls)* I am counting on your performance! By the way, the grade will be included in your report card grade.

Video 4 (explicit):

*Two female students walk from the blackboard to the back of their seats. The board says "Math Quiz". Teacher wipes away solutions of previous task and leaves heading "Math Quiz".*

*Teacher:* And we continue with our math quiz! Two students will compete against each other!  
Remember, speed counts! Julia and Jerome are next to go to the front of the blackboard.

*Julia and Jerome go to the blackboard. Teacher sets the task.*

*Teacher:* What are  $\frac{3}{4} * \frac{7}{4}$  ?

*Julia and Jerome calculate. Jerome writes:  $\frac{3}{4} * \frac{7}{4} = \frac{21}{16}$ . Julia writes  $\frac{3}{4} * \frac{7}{4}$  and then thinks about it longer.*

*Teacher:* Yes, okay, Julia, Jerome is already done. Jerome, that means an A for you. Julia, you weren't as fast as Jerome, but I'll give you the A too, but maybe with a little minus? Comparing the genders in math is simply unfair.

**Appendix B***Filler Items of the MGMQ-s*

1. Girls are more likely to dislike mathematics as a school subject than boys, because girls consider mathematics to be less relevant to them.
2. In girls and boys, similar brain regions are active during the processing of mathematical tasks.
3. As early as in elementary school, children think that mathematical professions are rather something for men and social professions are rather something for women.
4. On average, boys rate themselves as more competent than girls in terms of their mathematical abilities.
5. Girls would presumably be more motivated in mathematics if more topics with female connotations were the content of the tasks and lessons.
6. Boys identify more strongly with mathematics as a school subject than girls do.
7. In identity formation, girls often move away from the subject of mathematics in order to feel more like they belong to the female gender.
8. Even girls who have very good grades in mathematics rarely enter mathematical professions.

## Appendix C

### *Example Answers on the Video Analysis with Rating*

Video 1	Scoring (-3 to +4)	Reason for scoring with regard of awareness of math-gender stereotype- reinforcing behavior (MGSRB)
<p>Visualization makes sense, but it is incomplete. The question should also be written down on the board. Evaluation of the task as "good" and "we'll take it" is not motivating – rather the opposite! Gender cliché very problematic: Woman bakes a cakes / man is an engineer. Reproduces typical gender role models, demotivates girls in STEM.</p>	+4.0	<p>MGSRB is detected and critically evaluated with regard of gender-stereotypes and their consequences</p>
<p>Chalk-and-talk technique. Teacher reads aloud, students just listen, don't have the task available visually. Students are not involved in writing out the facts given. The only thing I would adopt [from the teacher's behavior] is that some of the information is written down for everyone. But [I would ask this to be done] by the students. The assignment also has to be read aloud by [one of] students. [Here, however:] students always just sit there passively. Also, the solution of the task should be provided by a student and written on the board – this should not be done by the teacher.</p>	0.0	<p>The MGSRB is not detected; the task is not focused; neither positively nor negatively evaluated</p>
<p>Lesson is teacher-centered. Students watch and the teacher explains on the board. The task: The second task is closer to everyday life than the first one. The students are passive. I would design the tasks so that the students have to figure out how to solve the task. However, the task is good; based on everyday life. Like shopping at the market I think is good. The way the task is presented here, you could also have used a purely numerical task.</p>	-1.0	<p>The task is perceived as good, without giving gender-stereotyped reasons. However, the task is in focus and the reproduction of gender-stereotypes is not detected or criticized.</p>

Video 2	Scoring (-3 to +4)	Reason for scoring with regard of awareness of MGSRB
<p>Using the random numbers generator to determine who should answer the question causes stress in students because it creates an exam situation. Therefore, I would not adopt this. After [the female student Lisa] answers the question, the teacher attempts praise. “Gee, great job”, however, he sounds like he didn’t think she [Lisa] could do it and conveys to the student [Lisa] that he thinks she can’t do math. This therefore does not build a positive culture of error, but encourages resignation (“He [the teacher] thinks I can’t do it anyway”, turns into [Lisa thinking] “I can’t do it anyway”) or fear of making mistakes because the student [Lisa] wants to prove to him [the teacher] that she can do it after all. The teacher again uses the random numbers generator to determine who corrects the answer: “What would be the correct way to say that?” Better to ask something like this: “Can you shorten this? If so name the shortened fraction.” The question the teacher asks here undermines confidence in mathematical ability in the [female] student [Lisa], (“Everyone else can do it better than I can”). The idea of showing the calculation using circles is good, but different shading for the different fractions would have been better for comprehension. The question to the student whose turn it was earlier [Lisa] embarrasses her. The task Steffen gets is much more difficult than the one Johanna gets. On the other hand, the praise Steffen gets for his correct answer is less euphoric than the one Johanna gets. This gives the impression that the teacher is assuming that boys can do math and girls can’t, thus promoting the social imbalance in this subject. The wording of the questions was also different: “Steffen, Name the smallest common...,” but “Johanna, can you...?” This calls into question Johanna’s ability to answer the question.</p>	+4.0	MGSRB is detected and critically evaluated with regard of gender-stereotypes and their consequences
[...] Teacher is appreciative and praises.	-1.0	Praise is perceived as good, regardless of gender stereotypical imbalance. Regardless of occupation with praise, no awareness of MGSRB.
<p>Strong preference for girls: They get easier tasks and stronger praise, boys are neglected. [...]</p>	-2.5	MGSRB is detected and criticized. Yet, no awareness for global problem of the math-gender stereotype. Instead fearing the teacher favoring the girls, which contradicts common research on the topic.

Video 3	Scoring (-3 to +4)	Reason for scoring with regard of awareness of MGSRB
<p>Unannounced tests are a great source of stress. This should be avoided if possible so that students can prepare and thus give their best. "Boys, I'm counting on you!" – This phrase shows unequal treatment of the genders in the classroom and reduces the learning opportunities of those who do not identify as boys. At the same time, it puts pressure on the boys. "By the way, the result count into the final grade", this sentence does not provide any new information, because everything that happens in class counts into the final grade, but it creates pressure and should therefore be avoided.</p>	+4.0	MGSRB is detected and critically evaluated with regard of gender-stereotypes and their consequences
<p>-I would not adopt the phrase: "This is a comparison test between all courses -I would adopt the phrase: "Boys, I'm counting on your performance."</p>	-1.5	No awareness of MGSRB. Gendered motivational phrase is perceived as good.
Video 4		
<p>A quiz sounds motivating at first. Bringing two students to the board allows a direct comparison of the learners' performance (who is faster, who has the right result) Teacher's comment [regarding comparison of the genders] sounds negatively pejorative and reinforces stereotypes about girls' poorer performance in math [math-gender stereotypes] – this behavior is demotivating for girls.</p>	+4.0	MGSRB is detected and critically evaluated with regard of gender-stereotypes and their consequences for female learners
<p>The math quiz can motivate and have a playful character, but it can also be stressful for some. To be considered: is it suitable for my learning group? The quiz should not be evaluated (with grades), because then it is no longer a quiz and loses the playful character. The comparison situation on the blackboard in front of everyone exposes weak students and should be avoided. It is better to ask only voluntary students to come to the blackboard! Do not grade the results. I would not do it like the colleague [in the video] in my lessons.</p>	0.0	No detection of MGSRB, no critical evaluation, no occupation with the problem in itself
<p>Playful situation in form of a quiz. Fair evaluation - both get an A for the correct result. The faster without minus, but this has no consequences. The teacher does very well here by explicitly not discriminating because of [the students'] gender. Good learning- / exam- or quiz atmosphere.</p>	-2.5	The MGSRB receives attention here, but is not recognized as such; no awareness of the problem and hence even praise for the (actually negative) MGSRB.

*Note.* MGSRB = math-gender stereotype-reinforcing behaviors.

**Appendix D***Follow-up Questions on the Video Analysis**C.1 Mental Effort*

The task (video analysis) was exhausting.

I made an effort in the task (video analysis).

*C.2 Relevant Model Teacher Characteristics*

To what extent did you notice the following characteristics of the teacher while watching and analyzing the instructional videos? (six-point Likert-scale from *not at all* to *very much*)

Ascribed ethnicity of the teacher

Age of the teacher

Attitude of the teacher

Gender of the teacher

*C.3 Questions on Technical Difficulties with the Videos*

I had technical difficulties with the videos.

(Dichotomous rating)

How was the technical quality of the videos?

(Six-point Likert-scale from *very bad* to *very good*)

Is there anything else you would like to let us know in relation to the videos?

(Open question)

-This manuscript is currently under review at Educational Psychology Review.-

**Should We Tell Teachers That They Are (In-)Correct?  
How Personalized Refutation Text Promotes or Hampers Conceptual Change**

Anna-Sophia Dersch<sup>1</sup>,


Ian Thacker <sup>2</sup>,


& Alexander Eitel<sup>1</sup>


Department of Psychology, Justus-Liebig-Universität Gießen<sup>1</sup>

Department of Educational Psychology, University of Texas at San Antonio<sup>2</sup>

**Author Note**

Anna-Sophia Dersch  <https://orcid.org/0000-0002-6012-3411>

Ian Thacker  <https://orcid.org/0000-0002-2492-2929>

Alexander Eitel  <https://orcid.org/0000-0001-6403-3467>

We have no conflicts of interests to disclose.

Correspondence concerning this article should be addressed to

Anna-Sophia Dersch,

Department of Psychology, Justus-Liebig-Universität Gießen.

Email: anna-sophia.dersch@psychol.uni-giessen.de

**Acknowledgments**

We would like to thank Prof. Dr. Stella Vosniadou for her expertise and advice during the process of writing this paper.

### Abstract

Teachers hold misconceptions that female learners have lower levels of mathematical talent than males. Such misconceptions are related to the math-gender stereotype that shapes how teachers interact with students and can contribute to the underrepresentation of women in mathematical fields. One method for changing these misconceptions is refutation text, which explicitly states and refutes a misconception and provides an alternate explanation. Since refutation text may not be sufficient for *teachers'* conceptual change, we applied personalized refutation text, which can successfully promote teachers' conceptual change. However, teachers may have strong attitudes related to math-gender, and refuting information about social-scientific topics, like math-gender misconceptions, might cause *worldview backfire effects* where teachers double down on their original stance.

In a randomized 2x2-pre-post-design with 336 teachers and student teachers, we investigated whether (1) expository texts, (2) personalized expository texts, (3) common refutation texts, or (3) personalized refutation texts were most effective for conceptual change. We found that personalized refutation texts initiated the strongest conceptual change for student teachers who had held misconceptions at pre-test. However, personalized refutation texts also triggered the most backfire among student teachers who had not held misconceptions at pre-test. Findings suggest that personalized refutation texts effectively catalyze teachers' conceptual change, but only if conceptual change is necessary. However, if misconceptions are not prevalent, personalized refutation text may lead to familiarity backfire; by repeatedly naming a misconception, it gains familiarity and becomes more credible than the intended refutation information. Hence, we recommend that personalized refutation interventions be adapted to individuals' prior knowledge.

## Introduction

Pre-service and in-service teachers have misconceptions about various psychological topics, like multimedia learning (Eitel et al., 2019; 2021), neurodevelopment (Gini et al., 2021), self-regulated learning (Vosniadou et al., 2020; 2021), and gendered ability differences, for instance, the misconception that female learners have lower mathematical talent (Dersch et al., 2022a). Teachers' misconceptions about women's mathematical ability can lead teachers to endorse stereotypes and communicate subtle messages to female students that impact their self-confidence and achievement (Carlana, 2019). Such misconceptions might also influence lesson design and teacher-student interactions, thus should be addressed (e.g., Authors, submitted; McAfee & Hofman, 2021; Prinz et al., 2021; Wilen, 2004). Refutation texts are, among other methods, one possible strategy to effectively change misconceptions (Tippett, 2010). Refutation texts name a commonly held misconception, state that it is incorrect and provide an alternative explanation (Tippett, 2010). However, common refutation texts are not always sufficient for teachers' sustainable conceptual change (Dersch et al., 2022b; Ferrero et al., 2020b; Zengilowski et al., 2021).

One factor that hinders teachers from revising their misconceptions is when they feel personally detached from the information presented in refutation texts—a problem that can be addressed through self-implication (Dersch et al., 2022b; Gregoire, 2003). Indeed, evidence suggests that *personalized feedback* that increases self-implication can be beneficial for reducing teachers' misconceptions about multimedia learning (Dersch et al., 2022b). Self-implication may, for instance, be expressed through negative affect related to internal blame, such as guilt or shame (Dersch et al., 2022b).

Yet, self-implication and consecutive negative affect can also be detrimental for refuting misconceptions, because it is related to *backfire effects* (Trevors et al., 2016)—which occurs when misconceptions are reinforced instead of reduced as a result from refutation information (e.g. Lewandowski, 2012). One set of topics that are prone to trigger backfire effects are social-scientific topics (Jacobson et al., 2012; Lewandowski, 2012; Sherman & Cohen, 2002). One social scientific topic that may trigger backfire effects are *math-gender misconceptions*—the incorrect belief that girls have lower mathematical talent than boys—misconceptions that are widely held by teachers (Dersch et al., 2022a).

To our knowledge, no research to date has studied whether math-gender misconceptions can be shifted with refutation texts, and whether personalized refutation texts predict more prominent shifts, or if they cause backfire effects. As such, in this study, we studied refuting teachers' math-gender misconceptions in a 2x2 experimental research design, using refutation texts vs. common expository texts, with personalized feedback vs. without personalized feedback. Namely, we investigated the benefits of personalizing refutation instruction while accounting for potential backfire effects.

## Theoretical Background

### Teachers' Math-gender Misconceptions

Misconceptions can be defined as knowledge conceptions that appear highly plausible while contradicting scientific evidence or established theory (Eitel et al., 2021; Hughes et al., 2013). Because knowledge conceptions are stored schematically in long-term memory (that is, integrated or intertwined with one-another; Sweller et al., 1998), misconceptions are robust and difficult to change. The more intertwined a misconception is with other knowledge conceptions, the harder it is to change them. This is consistent with theories on *conceptual change*—the changing of misconceptions that consider misconceptions to be: independent from other knowledge structures (e.g., diSessa, 1988; Dole & Sinatra, 1998) or organized into holistic theoretical frameworks (e.g., Vosniadou, 1994).

Two common math-gender misconceptions have been repeatedly found among student teachers and in-service teachers (e.g., Authors, submitted; Dersch et al., 2022a). One misconception is the *empathizing-systemizing* misconception (Dersch et al., 2022a). This misconception assumes that boys have higher inherent systematic thinking abilities due to pre-natal testosterone-exposure in comparison to girls who are assumed to have higher empathic thinking abilities. Higher systematic thinking in boys is assumed to lead to higher inherent mathematical talent in boys in comparison to girls (e.g., Baron-Cohen, 2005). This misconception assumes that empathic and systematic thinking are dualities, while in fact both are associated with general intelligence (g; Gottfredson, 1998; see also Martingano & Konrath, 2022). Additionally, this misconception assumes that systematic thinking leads to better math performance, when in fact they are not related (Escovar et al., 2016). Finally, this misconception is inconsistent with evidence that gender disparities in math performance are marginal to non-existent (e.g., Lindemann et al., 2010) and, if present, oftentimes are mediated by social factors like sex-role identity (Else-Quest et al., 2015). In this vein, Tiedemann (2002) found that teachers assumed that high-performing girls in math were less logical and exhibited less systematic thinking than high-performing boys.

Another common misconception is called *girls' compensation*. This misconception involves the belief that girls who earn similar math grades to boys do so because they are more diligent than boys, whereas boys achieve similar grades with less effort due to their innate talent (Dersch et al., 2022a). While research finds girls to be more diligent than boys across most subjects (e.g., Duckworth & Seligman, 2006), this diligence leads to on average *better* grades not similar grades—also in mathematics (Duckworth & Seligman, 2006). This misconception furthermore does not consider that most research concurs on girls not lacking inherent mathematical talent (Girelli, 2021; Kersey et al., 2019; Lachance & Mazzocco, 2006). Beliefs associated to this misconception could be identified in interviews with parents and teachers (Sáinz et al., 2020). Furthermore, in line

with this misconception, teachers attributed girls' medium-level mathematics performance to diligence and lack of talent, while they attributed similar medium-level math performance of boys to talent and lack of diligence (Tiedemann, 2002).

Both misconceptions were found to be prevalent among about a third of student teachers and in-service teachers across two different samples (Authors, submitted; Dersch et al., 2022a). These two misconceptions, just like other misconceptions, might negatively influence student-teacher interaction and lesson design (e.g., Prinz et al., 2021; Wilen, 2004). In this vein, the two misconceptions were found to be negatively related to awareness towards math-gender stereotype-reinforcing behaviors. Meaning, the more misconceptions in-service teachers held, the less likely they were to detect and critically evaluate a model teachers' gender stereotype-reinforcing behavior in a math lesson setting (Authors, submitted).

These two misconceptions—just like other misconceptions are related to *attitudes* (see also Jacobson et al., 2021), meaning, the positive or negative evaluation of an object or entity (Eagly & Chaiken, 1993). Specifically, math-gender misconceptions were found to be related to the math-gender stereotype (Authors, submitted; Dersch et al., 2022a). Stereotypes describe the cognitive component of an attitude and contain a set of correct- and incorrect beliefs about a social group based on their mere affiliation to said social group (e.g., Greenwald et al., 2002). The math-gender stereotype is a set of beliefs connoting mathematics as male (e.g., Ashmore & delBoca, 1979). According to dual process models, stereotypes are part of *heuristic processing* (e.g., Bodenhausen, 1990; Strack & Deutsch, 2015). Heuristic processing describes a shortcut within information processing based on consolidated, oversimplified or overgeneralized knowledge, reducing the sense of an environment's complexity (e.g., Fiske, 1982; Shah & Oppenheimer, 2008). The math-gender stereotype hence influences social perception, contributing to female learners' lower interest, motivation and performance in mathematics (Cvencek et al., 2012; Greenwald et al., 2002; Girelli, 2023; Steele, 2003), and eventually to low female representation in mathematics, meaning, less women pursuing careers in the field of mathematics (Eccles, 2011; Hand et al., 2017; Halpern et al., 2007; Nosek & Symth, 2011). Teachers, on average, hold the math-gender stereotype (Authors, submitted; Dersch et al., 2022a; Eccles, 2011; Gunderson et al., 2012). Furthermore, teachers are important socializers and are likely to contribute to gender disparities in math by manifesting the math-gender stereotype in their instruction and interactions with students (e.g., Eccles, 2011; Gunderson et al., 2012). Hence, the math-gender stereotype should be reduced among teachers. However, (math-gender) stereotypes have rarely been successfully and sustainably reduced (FitzGerald et al., 2019; Kollmayer et al., 2020). That is likely because they are global cognitive structures (Greenwald et al., 2002), and therefore hard to change (Dersch et al., 2022a). Misconceptions, on the other hand are more specific and hence a good starting point to conceptual change (Larkin,

2012) and prospective attitude change (Jacobson et al., 2021). Due to the prevalence of math-gender misconceptions among teachers (Authors, submitted; Dersch et al., 2022a), and their relation with the math-gender stereotype (Dersch et al., 2022a; Girelli, 2022), math-gender misconceptions should be changed in student teachers and in-service teachers as important (future) socializers.

### **Changing Teachers' Math-gender Misconceptions Using Refutation Text**

While some theoretical approaches differentiate between the concepts of conceptual change and knowledge revision (Kendeou, 2024), other approaches define knowledge revision as one facet within conceptual change (Chi, 2009) or use both terms, conceptual change and knowledge revision, synonymously (Mason, 2009). In the following we will also use the terms conceptual change and knowledge revision synonymously. For theoretical frameworks, we will consult the Knowledge Revision Components Framework (KReC) by Kendeou and O'Brien (2014), which is supported by strong evidence in the subject of revising misconceptions (Danielson, 2024) as well as the Cognitive Affective Model of Conceptual Change (CAMCC; Gregoire, 2003), which is suited especially for teachers' conceptual change (Gregoire, 2003).

The KReC (Kendeou & O'Brien, 2014) includes three processes necessary for conceptual change. According to the KReC, conceptual change occurs through (1) *co-activation* of the misconception and the correct information that is aimed to refute the misconception. Co-activation can for instance be achieved with refutation texts (Danielson et al., 2024). Refutation texts present a commonly held misconception (e.g., for the math-gender misconception of *empathizing-systemizing*: "Many people think that boys inherently think more systematically than girls. Due to systematic thinking being assumed necessary for math, many people assume that boys are better at math from birth."). Then refutation texts state that it is incorrect and provide an alternative scientific explanation supporting the correct conception (Tippett, 2010). Hence, the named misconception and the provided alternative explanation are co-activated. (2) The second process according to the KReC is *integration* of the correct information in the relevant cognitive schema or construct. Co-activation due to a refutation text should trigger integration within the relevant knowledge framework or theory (see also diSessa, 1988; Vosniadou, 1994), for math-gender misconceptions within the math-gender stereotype and other related knowledge. (3) The third process is *competing-activation*, which describes that the misconception remains within the cognitive schema but the correct information is activated simultaneously—iteratively stronger than the misconception, while the misconception is inhibited in relevant situations (Kendeou & O'Brien, 2014). Like this, the misconceptions becomes less relevant and the correct conception gains relevance. Co-activation and integration facilitated by refutation texts should prospectively foster competing-activation, which is the goal of knowledge revision (Kendeou et al., 2013; 2014).

This elaborate process of conceptual change, according to dual-processing theories (e.g., Maheswaran & Chaiken, 1991), the Cognitive-Affective Model of Conceptual Change (CAMCC; Gregoire, 2003) and persuasion- and attitude change theories (e.g., Nabi, 1999) requires deliberate systematic processing. Since systematic processing requires active mental effort (Klepsch & Seufert, 2021), motivation to actively engage with the presented information is necessary (Eagly & Chaiken, 1993; Nabi, 1999). By highlighting the threshold between one's own potential misconception ("Many people believe that...), the scientifically correct conception ("However, that is incorrect."), and presenting an incongruent message, refutation texts have the potential to increase self-implication, meaning a learners' understanding, that they, themselves, hold a misconceptions, feeling in charge to change it (Gregoire, 2003). Such self-implication can increase motivation for systematic processing of refutational messages (e.g., Maheswaran & Chaiken, 1991) and therefore for conceptual change (Sinatra, 2005; cf. Gregoire, 2003). One set of factors that drive increased motivation for change are negative emotions that result from internal blame, such as feelings of guilt or shame (Dersch et al., 2022b; Gregoire, 2003; Nabi, 1999; Trevors & Kendeou, 2020). Internal blame happens when learners take ownership of their misconception and is expressed as feelings of guilt or shame, which can be contrasted with situations where learners blame external factors for misconceptions, and is expressed as anger (cf., internal/external dimensions of attribution; Weiner, 1985; and internal/external locus of control; Lefcourt, 1991).

The CAMCC (Gregoire, 2003) posits that *teachers'* conceptual change is especially challenging to achieve because teachers have experienced school during their primary and secondary education, which prior to their teacher education has already consolidated their knowledge, beliefs—and therefore also their misconceptions—about school, learning and teaching (Gregoire, 2003). Additionally, teachers may not feel self-implicated when confronted with misconceptions about teaching and learning considering themselves knowledgeable on the topic (Dersch et al., 2022b). Teachers may think, "Many people hold this misconception, but I am not one of those" (see also Dersch et al., 2022b) and process the refutational information heuristically (Gregoire, 2003). Hence, for teachers, reading a refutation text may not increase self-implication enough to process the refutational information systematically instead of heuristically. Another potential reason for such heuristic processing hampering conceptual change is the topic of the targeted misconceptions. Teachers' math-gender misconceptions are related to and likely embedded in the belief system of the math-gender stereotype (Authors, submitted; Dersch et al., 2022a; see also von Aufschnaiter & Rogge, 2015). As mentioned above, the math-gender stereotype is part of heuristic processing (Bodenhausen, 1990; Strack & Deutsch, 2015), which, in contrast to systematic processing, relies on cognitive shortcuts, and, instead of actively constructing new knowledge or restructuring pre-existing incorrect knowledge, draws on prior, often oversimplified or overgeneralized knowledge

(Greenwald, 2002). Aiming to refute math-gender misconceptions may activate the math-gender stereotype, which may activate heuristic processing, leading teachers to draw on oversimplified prior knowledge about mathematics and gender. This process might endanger systematic processing of the refutational information, which would, be required for active restructuring of knowledge and conceptual change (Gregoire, 2003). Hence the activation of the math-gender stereotype when targeting math-gender misconceptions may hamper conceptual change.

To date, most research on student teachers' and in-service teachers' conceptual change focuses on learning and the brain (e.g., Ferrero et al., 2020a; 2020b; Grospietsch & Mayer, 2018), or more specifically on multimedia learning (e.g., Dersch et al., 2022b; Prinz et al., 2021). On the topic of multimedia learning, findings suggest that refutation texts effectively foster teachers' conceptual change in comparison to no intervention (Ferrero et al., 2020a). However, the effects were not stable over time (Ferrero et al., 2020a; 2020b). Furthermore, most research did not find that refutation text influenced behavior or transfer outcomes (Dersch et al., 2022b; Ferrero et al., 2020a; 2020b; Prinz et al., 2021). Lastly, the common refutation text did not outperform expository text in an experiment, likely due to lacking self-implication when reading the refutation texts (Dersch et al., 2022b).

In line with this finding, the CAMCC (Gregoire, 2003) posits that personal relevance or self-implication is necessary to avoid heuristic processing and instead foster systematic processing of refutational information, which is a necessary factor for successful conceptual change (see also Dole & Sinatra, 1998). One factor increasing self-implication for conceptual change may be *personalized feedback* in refutation texts (Dersch et al., 2022b).

### **Personalized Refutation Texts**

Personalized messages with specific feedback (e.g., "your answers indicate that..."), can lead to higher personal involvement, and therefore to more systematic and focused information processing (Mayer, 2014; Renkl, 2015). Most research on *personalized feedback* focuses on learners' performance in a prior knowledge test (Maier & Klotz, 2022). Such personalized feedback has the potential to improve learners' performance (Narciss et al., 2014). Personalized feedback in a refutation text aims at creating an impasse experience, highlighting the difference between one's own knowledge (or potential misconception) and the correct information (Dersch et al., 2022b; Sánchez et al., 2009). Such an impasse experience can, combined with a refutation text, further increase self-implication, which helps learners understand that they themselves hold the misconception and need to change it (Dersch et al., 2022b). Furthermore, refutation texts have been critiqued for being impersonal and for not taking learners' individual prerequisites, such as prior knowledge, into account (Asterhan & Dotan, 2018; Zengilowski et al., 2021). Combining refutation texts with personalized feedback, however, combines the strengths of both methods. Accordingly, personalized refutation interventions have been found to be beneficial for conceptual change (Asterhan & Dotan, 2018), especially among teachers (Dersch et al., 2022b).

In a study by Dersch and colleagues (2022b), a common refutation text was found to be as effective as an expository text for promoting in-service teachers' conceptual change about multimedia learning. However, conceptual change improved when the refutation text was combined with personalized feedback that shared information about the teachers' misconception. The effect was mediated by feelings of internal blame (i.e., guilt and shame; Dersch et al., 2022b). While targeting these teachers' misconceptions about multimedia learning led to an increase in negative affect related to *internal* blame, it was not related to an increase in negative affect related to *external* blame—with both in-service teachers (Dersch et al., 2022b) and student teachers (Prinz et al., 2021). Such negative affect of external blame would have been a potential symptom of a threat appraisal of a refutational message (Dersch et al., 2022b; Gregoire, 2003; Trevors et al., 2016). Negative affect related to external blame may be an indicator for repulse (Dersch et al., 2022b; Festinger et al., 1962). Repulse has been found to hinder learning (Chinn & Brewer, 1993), for instance due to the discounting of the refutation message leading to superficial, heuristic processing (e.g., Gregoire, 2003). According to the CAMCC (Gregoire, 2003), threat appraisal when confronted with refutational information is more likely or stronger when self-implication is increased. Hence, while personalized refutation texts hold potential for promoting in-service and preservice teachers' conceptual change, they also hold risks. In comparison to non-personalized refutation texts, personalized refutation texts can potentially lead to stronger repulse, hampered learning (Gregoire, 2003), or even *backfire effects* (Lewandowsky & Cook, 2011; Jacobson et al., 2021; Trevors et al., 2016).

### **Backfire Effects**

Backfire effects are the ironic reinforcement of a misconception after an attempted refutation (Lewandowsky et al., 2012). Backfire effects can, for instance, occur due to a refutational message contradicting learners' core values or endangering learners' self-concept, meaning, the combination of all schemas a learner holds about themselves and potentially want to maintain (Lewandowsky et al., 2020; Muis et al., 2015; Sherman & Cohen, 2002; Trevors et al., 2016). Backfire effects triggered by a potential endangerment of the self-concept are specified as *worldview backfire* (Lewandowsky et al., 2020). Worldview backfire can be explained by *passive activation*. Passive activation describes the mechanism of related constructs such as prior knowledge but also attitudes and emotions being unintentionally co-activated, when a certain (mis-)conception is activated (Ecker et al., 2014) for example when reading a refutation text (Kendeou & O'Brien, 2014). Whenever a misconception is related to attitudes, this increases the likelihood of those attitudes being activated when reading a refutation text about said misconception. If those attitudes are related to the self-concept, or core values, which an attempted refutation endangers, this increases the likelihood of negative affect of external blame and repulse, expressed for

instance through anger, which is related to backfire effects (Gregoire, 2003; Sherman & Cohen, 2002; Trevors et al., 2016).

While there have been some failed attempts to replicate backfire effects in recent research (e.g., Ecker et al., 2023), worldview backfire is still the most likely backfire effect to occur (Lewandowsky et al., 2020), specifically for social-scientific topics (Jacobson et al., 2021). Such topics are oftentimes related to attitudes, core values or learners' self-concept (Sherman & Cohen, 2002; Trevors et al., 2016). Math-gender misconceptions are a social-scientific topic combining knowledge about mathematics and gender. Mathematics is related to self-concept, since mathematics is an important skill in education and beyond. Knowledge about one's own mathematical competence, perceived difficulty of mathematics and beliefs about what is necessary to be good at mathematics hence forms part of the ability self-concept, or the academic self-concept, meaning implicit and explicit knowledge beliefs about (one's own) academic abilities (Arens et al., 2021; Brookover et al., 1964; Ommundsen et al., 2007). Gender also forms part of the self-concept (Crain, 1996). Gender is an important social characteristic (Abrams et al., 1990), and an important factor for socialization in most societies (Stockard, 2006). Ideas about gendered mathematical ability differences are furthermore likely to have influenced many people's own education and contributed to their career choices (e.g., Eccles, 2011; Wang & Degol, 2017). Hence, aiming to refute *math-gender* misconceptions might endanger the self-concept in different ways. This endangerment might then trigger worldview backfire.

Furthermore, backfire, according to spreading activation, may not only reinforce the misconceptions themselves but also the related negative attitudes (Jacobson et al., 2021), like the math-gender stereotype. Jacobson et al. (2021) hence distinguish *knowledge backfire* and *attitude backfire*. Knowledge backfire describes the ironic reinforcement of a misconception itself, whereas attitude backfire describes the ironic reinforcement of a related attitude—or of a stereotype when aimed to be refuted or changed (Jacobson et al., 2021).

When already expecting knowledge and attitude backfire, for a social-scientific topic like math-gender misconceptions, personalized refutation texts further increasing self-implication might also increase perceived stress caused by a message conflicting with own knowledge and beliefs (Gregoire, 2003), leading to stronger threat appraisal, more repulse (Gregoire, 2003) and a potentially even stronger backfire effect. We sought to examine whether this would be the case for math-gender misconceptions.

### **Current Study and Hypotheses**

In this study, we aim at adding to the body of research on the refutation paradigm, which has been critiqued for focusing on few refutation text topics (Zengilowski et al., 2021), by testing it for the new topic of math-gender misconceptions. The refutation text paradigm has furthermore been critiqued for being overly

authoritative (Zengilowski et al., 2021) and not addressing learners' personal prerequisites (Asterhan & Dotan, 2018). Hence, in this study we address this by researching the effects of *personalized* refutation texts (= personalized feedback + refutation text). Specifically, we research whether personalized feedback with expository text can be sufficient for teachers' conceptual change or if personalized refutation text remains the strongest method for teachers' conceptual change. We also include the behavioral measure of teachers' awareness towards math-gender stereotype-reinforcing behaviors to test the impact of our intervention on teachers' perception of relevant situations (Authors, submitted). We do so by applying a 2x2 design with refutation text vs. expository text, and personalization vs. no personalization. We further want to find out if (personalized) refutation text can hamper conceptual change on the social-scientific topic of math-gender misconceptions, which might cause backfire effects. Additionally, we research the role of negative affect of internal blame in driving conceptual change and negative affect of external blame hampering conceptual change or causing backfire effects. Lastly, we want to explore a potential influence on the math-gender stereotype and specific attitudes towards closing the math-gender gap. Hence, we aimed to address the following preregistered hypotheses ([https://aspredicted.org/F3R\\_X79](https://aspredicted.org/F3R_X79)).<sup>3</sup>

## **Main Hypotheses**

### ***H1. Personalized Refutation Text Hypothesis***

We predict that the positive effect of personalized refutation texts on teachers' conceptual change is stable for a social-scientific topic, which is more likely to trigger backfire effects (e.g., Trevors et al., 2016). Teachers' conceptual change, meaning their reduction of math-gender misconceptions, should be strongest when learning from personalized refutation texts. Furthermore, teachers should be most aware towards math-gender stereotype-reinforcing behaviors in this condition.

### ***H2. Shame-drives-change Hypothesis***

We assume that negative affect related to internal blame mediates the positive effect of personalized refutation text on conceptual change and awareness towards math-gender stereotype-reinforcing behaviors. Participants who learn about math-gender misconceptions from personalized refutation text should experience more negative affect related to internal blame. This increased negative affect should increase conceptual change.

---

<sup>3</sup> In this first experiment, we decided to focus on studying the hypothesis regarding conceptual change. Based on our experimental design, we can only draw preliminary conclusions on potential attitude change. Hence, we decided to change the order of the preregistered hypotheses and adapt their wording accordingly. While the attitude-hypotheses allow for a first exploration, further research is necessary, to assess a potential impact on attitude more in detail.

### ***H3. Knowledge Backfire Hypotheses***

Contrary to our prior hypotheses, and also based on our literature review, the interventions could also cause backfire effects. Conceptual change effects might then be on average less pronounced or misconceptions might even be reinforced. This may be likely for personalized and non-personalized refutation texts. In these conditions, awareness towards math-gender stereotype-reinforcing behaviors would then be lowest in comparison to the other conditions.

### ***H4. Anger-mediates-knowledge-backfire Hypothesis***

Knowledge backfire might be mediated by negative affect related to external blame. Participants who read (personalized) refutation texts might experience more negative affect related to external blame (e.g., anger). This increased affect of external blame might lead to less awareness towards math-gender stereotype-reinforcing behaviors or to less conceptual change and misconceptions being reinforced among these participants in comparison to those who learn from (non-personalized) expository texts.

### **Exploratory Hypotheses**

#### ***H5. Conceptual-change-makes-attitude-change Hypothesis***

We explore if reducing math-gender misconceptions reduces the math-gender stereotype. The most effective condition at fostering conceptual change might reduce the math-gender stereotype most.

#### ***H6. Attitude Backfire Hypothesis***

The interventions might further cause attitude backfire for the topic of math-gender misconceptions / the math-gender stereotype. Attitudes might then be on average more negative in a condition with stronger backfire effects. This might be likely for personalized and non-personalized refutation texts.

#### ***H5b. Anger-mediates-attitude-backfire Hypothesis***

Potential attitude backfire might also be mediated by negative affect related to external blame (e.g., anger). Participants who receive (personalized) refutation texts might experience more negative affect related to external blame. This increased affect of external blame might negatively affect the attitude of these participants in comparison to participants reading (non-personalized) expository texts.

### **Method**

The local ethics board approved this study (LEK FB06 2020-0042).

All research data has been made publicly available, including syntax for statistical analyses, via [https://osf.io/tp2sn/?view\\_only=4b1e6050d2d74e57b91c377742be973e](https://osf.io/tp2sn/?view_only=4b1e6050d2d74e57b91c377742be973e) (link will be converted into a doi upon publication of this manuscript).

## Participants and Design

Participants were recruited online via a university student recruiting system in Germany (SONA) and an online news page for teachers in German language (News4Teachers). Participants were screened to be either student teachers or in-service teachers prior to participation. Initially, 476 teachers provided informed consent to our online study. Of these, 42 participants dropped out before starting the study, and 84 participants dropped out over the course of the study. Dropout was similarly distributed among conditions,  $\chi^2(3) = 5.22, p = .156$ .

Overall, 350 participants completed the online learning experiments. Two of these 350 participants were excluded because they indicated that they had not participated in the study faithfully, another 12 were excluded because they indicated they had been so distracted during their participation so that their data should not be used, yielding a final sample of  $N = 336$  participants (266 female, 69 male, 1 non-binary;  $M_{\text{age}} = 27.56 [SD = 12.77, \text{range} = 18\text{--}75 \text{ years}]$ ).<sup>4</sup> Participating (student) teachers had on average 5.08 years of teaching experience ( $SD = 9.52$ ). Of these participating (student) teachers, 85 (25.3 %) were fully examined in-service teachers, 50 (14.9 %) were student teachers advanced in their training who already taught at schools, and 201 (59.8 %) indicated that they did not have any teaching experience, yet. A majority of the participating teachers ( $n = 204; 60.7 \%$ ) taught or studied to teach at secondary schools, while the remaining participants ( $n = 132; 39.3 \%$ ) taught or studied to teach at primary schools. For a detailed overview of the distribution of these demographics among conditions, see Table 1.

We had initially preregistered a sample of  $N = 250$ . We aimed at this sample consisting of substantial subsamples of teachers and student teachers. However, when reaching 250 participants, only few in-service teachers had participated, which is why we collected more data. Yielding a subsample of  $n = 85$  teachers and  $n = 251$  student teachers.

Participants were randomly assigned to one of four experimental conditions in a 2x2 design (personalized feedback and refutation text [ $n = 93$ ]; personalized feedback and expository text [ $n = 86$ ]; common refutation text [ $n = 74$ ]; common expository text [ $n = 83$ ]).

## Materials and Experimental Manipulations

For a graphic depiction of the manipulation described in the following, see Figure 1. In all conditions, participants read texts about mathematics and gender. The texts matched the misconception scales from the pre-test. One text targeted the *empathizing-systemizing misconception* (Dersch et al., 2022a). In all four conditions, the text started with the question “Do boys think more systematically, whereas girls think more empathically—

---

<sup>4</sup> Of these 336 participants, 266 participants completed the video task (for description, see Materials section). Since the video task consisted of four short videos, 70 participants had to be excluded due to technical difficulties (e.g., no headphones at hand in a public area, technical issues with playing the videos fluently due to bad internet connection).

and what influence does this have on mathematical talent?” Then the text presented an explanation with scientific evidence on the topic (163 words). The second text was targeted at the *girls’ compensation misconception* (Dersch et al., 2022a). In all four conditions, the text started with the question “Do girls compensate for their lack of mathematical talent through harder work?” Then the text presented an explanation with scientific evidence on the topic (190 words). The common expository text was the same for all for conditions and contained scientific references in APA-style, which aims at increasing text credibility (e.g., Vaughn, 2018). The other conditions (personalized refutation text, personalized expository text, refutation text) contained additional text components, which will be described next.

In the personalized text conditions (personalized refutation text, personalized expository text), participants additionally received a feedback based on their prior performance in the pre-test (43 – 52 words). Feedback on performance in the *empathizing-systemizing* scale was provided before the text about empathizing-systemizing, feedback on performance in the *girls’ compensation* scale was provided before the text about girls’ compensation. Feedback reflected participants’ answers. This means participants were either informed that their answers were (1) in line with current scientific evidence (i.e., they had correctly disagreed with all items per misconception scale), (2) reflected uncertainty (i.e., they had agreed to at least one misconception item per scale, while relatively uncertain [*absolutely uncertain* to *somewhat certain*]) or (3) were not in line with current scientific evidence (i.e., they had agreed to at least one item per misconception scale with high certainty [*certain* or *absolutely certain*]; cf. Dersch et al., 2022b). As an example, the feedback before the empathizing-systemizing text if participants had correctly disagreed with all misconception items on this scale was:

The evaluation of your answers shows that you have answered the questions on empathic and systematic thinking in boys and girls and their effects on mathematical talent in line with the scientific findings. In the following you can familiarize yourself once again with the scientific findings and supplement your knowledge on the subject.

In the common text conditions (common refutation text or common expository text) participants did not receive personalized feedback on prior performance. In the refutation text conditions (common refutation text, personalized refutation text), before reading the expository texts (and after potential personalized feedback in the personalized refutation text condition), participants read a refutation (32—77 words). For the empathizing-systemizing text, this refutation was

Many people think that the male and female brains are different from birth and that boys therefore think more systematically than girls from birth. According to this conceptions, systematic thinking is assumed to be fundamental for mastering mathematics. On the other hand, girls are assumed to think more

empathically, which according to this assumption would be less helpful for mastering mathematics. Accordingly, many people assume that boys are better at math from birth. However, these assumptions are incorrect and not in line with the scientific consensus.

Participants in the expository text conditions (personalized expository text or common expository text) did not receive this refutation. All texts were in German. According to Flesch-Kincaid index for German language texts, all texts were at a college student level (30 – 44). This is in line with the target group for this intervention being university students and graduated in-service teachers. For a translation into English of all text types (personalized refutation text, personalized expository text, refutation text, expository text) about both topics (empathizing-systemizing, girls' compensation), see Appendix A.

## Measures

### *Short Version of the Math-gender Misconception Questionnaire*

For the pre-test and the post-test, we assessed teachers' misconceptions with the short version of the Math-gender Misconceptions Questionnaire (MGMQ-s; Authors, submitted). The MGMQ-s consists of two misconceptions scales (*empathizing-systemizing* [e.g., "As girls think rather empathically and boys think rather systematically, boys are on average more talented in mathematics than girls.";  $\omega = .88$ ] and *girls' compensation* [e.g., "For the same grades in mathematics, boys, on average, work less hard than girls, as boys have more natural talent.";  $\omega = .70$ ]; Authors, submitted). The internal consistencies of both scales were similar in our sample (see Table 2). Both scales contain four items each. Since these eight misconception items describe scientifically unsupported statements, the correct answer is disagreeing with them. Hence, the MGMQ-s additionally contains eight (50 %) correct filler items to balance the questionnaire and avoid answer tendencies (Moosbrugger & Kevala, 2012; cf. Dersch et al., 2022a; Eitel et al., 2021). Filler items describe *true* statements related to the math-gender gap. Filler items were not analyzed here. To control for positioning effects, all MGMQ-s items were presented in a newly randomized order for each participant (cf. Dersch et al., 2022a).

For each item, participants first dichotomously indicated if they thought the statement was *correct* or *incorrect* (i.e., agreement). Then, participants indicated their *certainty* regarding the (in-)correctness of said statement on a five-point Likert-scale from *absolutely uncertain* to *absolutely certain* (i.e., certainty; cf. Dersch et al., 2022a; Eitel et al., 2021). We then calculated misconception scores with the four misconception items per scale by multiplying agreement (coded with +1) or disagreement (coded with -1) and response certainty (coded from 0 = *very uncertain* to 4 = *very certain*). Stronger misconceptions were hence expressed through higher misconception scores (see Table 2, for descriptive values). The range of possible values per item were -4 to +4 (cf. Authors, submitted; Dersch et al., 2022a; Eitel et al., 2021). We computed an average misconception score from both

misconceptions, both for the pre-test and the post-test. *Conceptual change scores* were then calculated by subtracting the pre-test result from the post-test result. Additionally, we used a criterion to determine whether a misconception was present or not. A misconception was assumed to be prevalent, if a participant agreed to at least one item per the respective misconception scale with high certainty (*certain* or *absolutely certain*; cf. Dersch et al., 2022b).

### ***Affect and Cognitive Processes During the Intervention***

To measure participants' affect, after learning from the texts we used the Positive- and Negative Affect Schedule (PANAS) for the moment (Watson et al., 1988), which measures the affect in the present moment. It consists of 20 items. Each item is one affect word (e.g., *guilty*). Participants rated the intensity of each affect on a five-point Likert-scale from '*little or not at all*' to '*extremely*'. We were interested in the negative affect scale. The internal consistency for the negative affect scale is  $\alpha_{\text{Cronbach}} = .89$  (good; Watson et al., 1988).

We further assessed participants' extraneous and intrinsic cognitive load using the two respective scales of the Naïve Cognitive Load Questionnaire (Klepsch et al., 2017) to control for potential differences in text complexity between conditions, which we assumed to be expressed in differences in extraneous or intrinsic load (see also Dersch et al., 2022b). We adapted the items to better fit our intervention being reading texts. All items have 7-point Likert scales from *completely wrong* to *absolutely right*. We used all three items for extraneous load (e.g., "The design of the texts was very inconvenient for learning"). The internal consistency of the adapted extraneous load scale was good,  $\alpha_{\text{Cronbach}} = .82$ . We used both items for intrinsic load (e.g., "Reading the texts was very complex."). The internal consistency of the adapted intrinsic load scale was acceptable,  $\alpha_{\text{Cronbach}} = .71$ . We further assessed participants' active and passive mental effort during reading the texts (Klepsch & Seufert, 2021). These items were also adapted to the reading task (e.g., "I put effort into reading the texts.").

### ***Behavioral Measure***

We included a behavioral measure, which in a prior study had been correlated with math-gender misconceptions (Authors, submitted), to determine if our intervention would not only influence conceptual change but potentially also influence behaviors. The behavioral measure was a video task testing awareness towards math-gender stereotype-reinforcing behaviors. Before watching the videos, participants read instructions stating that the following task was independent from the prior tasks about mathematics and gender and that we wanted to assess their general teaching abilities. Since in a real-life situation math-gender stereotype-reinforcing behaviors would occur without prior warning or learning about the topic, we aimed at increasing ecological validity like this. Hence, participants did not know that such behaviors would occur (see also Authors, submitted).

The videos were fully scripted. Each classroom video showed a female model teacher's math-gender stereotype-reinforcing behavior in a math lesson, for example the teacher praising girls exaggeratedly for very easy tasks while praising boys moderately for more difficult tasks. Furthermore, the videos contained distractors, for example the female model teacher using a random numbers generator. For all full video scripts, see Authors (submitted).

The videos were presented each on a separate page. The videos were between 0:12 minutes and 2:10 minutes. The videos were presented in the same order for all participants because the math-gender stereotype-reinforcing behaviors increased in explicitness between videos. *During* video presentation participants answer the open question: "List any relevant teaching strategies and teacher behaviors you notice. Which of these behaviors and strategies would you adopt and which not? Give a brief explanation." Participants could watch each video once, like in a real-life situation. However, participants could pause the video as often as they wanted to, to prevent them from experiencing time constraints while answering the question for each video (Authors, submitted).

We combined teachers' *detection* of math-gender stereotype-reinforcing behaviors (1 = mentioning; 0 = not mentioning) and participants' *critical evaluation* of math-gender stereotype-reinforcing behaviors (meaning if they argued with the math-gender gap, demotivation of girls and this being possibly detrimental; -3 to +3) by summing them up to *awareness*. For the full scoring rationale, see Authors (submitted). Two independent raters quantified participants' open answers. One rater had a degree in psychology and substantial knowledge about the math-gender stereotype, doing their PhD on the topic, the other rater was a student teacher in their final year before first state examination and trained by the first rater. Both raters were blinded for conditions. The first rater quantified 100 % of all answers, while the second rater quantified about 50 % to assess interrater reliability. Interrater reliability was good,  $\kappa_{\text{Cohen}} = .73$ ,  $z = 13.43$ ,  $p < .001$  (Altmann, 1990).

Here, awareness towards math-gender stereotype-reinforcing behaviors did not correlate with misconceptions in the pre-test,  $r = -.03$ ,  $p = .589$ , nor with misconceptions in the post-test,  $r = -.08$ ,  $p = .173$ , nor with conceptual change,  $r = .03$ ,  $p = .585$ . Hence, we treat it as additional behavioral measure, not as a transfer measure.

### ***Attitude and Stereotype Measures***

Participants indicated how female-to-male they perceived math from 1 (*very female*) via 5 (*neutral*) to 9 (*very male*) on a rating scale (Dersch et al., 2022a; Nosek et al., 2010; Steffens & Jelenec, 2011).

Furthermore, participants answered two items about the math-gender gap ("The math-gender gap, meaning women being less represented in the field of mathematics than men, should be reduced.", "Equal

treatment and rights of the genders in mathematical jobs is important.”;  $r = .34$ ), each on a five-point Likert-scale from *do not agree at all* to *fully agree*.

### **Demographics**

We asked for participants' age, their sex assigned at birth, gender they identify with, native language, teaching (or study) subjects, school type they will work at or already work at, and semesters studied as well as their practical experience in teaching.

### **Control for Faithful Participation**

At the end of the survey, participants answered two dichotomous questions (“Did you follow the instructions while completing this study?”, “Were you able to complete the questionnaire without distraction?”). Being an online experiment, these items allowed for an assessment of data quality. Participants who did not follow the instructions ( $n = 2$ ) or participants who reported they had been too distracted during participation for their data to be used ( $n = 13$ ) were excluded from further analysis.

### **Procedure**

The experiment was conducted in an online setting (Questback by Unipark). Before starting the study, participants gave informed consent. Then participants filled in the MGMQ-s (Authors, submitted) as pre-test. In the following learning phase, all participants read two texts about mathematics and gender (targeted at the empathizing-systemizing misconception and girls' compensation misconception). The instruction for the learning phase was the same for all four conditions: “Please read the two following texts about mathematics and gender differences thoroughly.” Text style depended on the condition participants were in (common refutation text, common expository text, personalized refutation texts, personalized expository text). Participants could only forward the page after 35 seconds per text, which is the time to quickly read each text. Hence, the learning phase took minimum 70 seconds. After the learning phase, participants answered questions about affect, cognitive load, and mental effort. Then, participants took the MGMQ-s again as a post-test. This post-test was followed by the Video Task. Afterwards, participants indicated their math-gender stereotype, their feminism and their attitudes towards mathematics and gender. Participants then answered questions about their demographics and their thorough participation. Finally, participants were fully debriefed and thanked for their participation. The participation in this online experiment took on average  $Md = 29,14$  minutes.

### **Results**

We conducted preliminary analyses, before testing the hypotheses. We used IBM SPSS statistics v30 for all analyses. For our hypotheses, we applied orthogonal contrast analyses, an ANOVA (both at an  $\alpha$ -level of .05) and mediation analyses with the PROCESS macro for SPSS (10,000 bootstrapping samples; Hayes, 2018). For

effect sizes we reported partial eta-squared and R-squared. Small values are .01, medium values are .06 and larger values are .14.

### **Preliminary Analyses**

According to histograms, both dependent variables, conceptual change and awareness towards math-gender stereotype-reinforcing behaviors, were approximately normally distributed in each condition. According to  $\chi^2$ -tests, gender, German as native language, school type to teach at (primary or secondary school) and status (student teacher vs. fully examined in-service teacher) were equally distributed among all four conditions ( $ps > .05$ ). According to ANOVAs, age, teaching experience, and misconceptions in the pre-test did not differ between all four conditions ( $ps > .05$ ). These results suggest a successful randomization. Intrinsic cognitive load and extraneous cognitive load also did not differ between conditions ( $ps > .05$ ). These results suggest that varying text length (personalized refutation text > refutation text  $\approx$  personalized expository text > expository text) did not influence intrinsic or extraneous cognitive load.

For an overview of the distribution of control variables, see Table 1.

### **Preregistered Main Analyses**

#### ***H1. Personalized Refutation Text Hypothesis***

**Conceptual Change.** Teachers' conceptual change should be strongest and hence conceptual change scores highest for those who read personalized refutation texts. We used orthogonal contrast analyses to test this pattern of one condition being stronger than the other three conditions in a 2x2 design. We translated the hypothesis into a contrast, in which common expository text, common refutation text and personalized expository text conditions were supposed to display lower conceptual change scores than the condition with personalized refutation text (contrast #1: [-1 -1 -1 +3]). To rule out possible alternative explanations, within the second contrast, we tested if personalized expository text differed from the other two conditions, meaning if personalized refutation text was sufficient to foster teachers' conceptual change (contrast #2: [-1 -1 +2 0]). Within the third contrast, we tested if common expository text and common refutation text differed from one another (contrast #3: [-1 +1 0 0]), to rule out this alternative explanation. The first contrast as our assumed explanation and a combination of the alternative explanations of the second and third contrast served as independent variables in a linear regression analysis with conceptual change score (post-test result – pre-test result) as dependent variable. In a regression model with all 336 participants, we did not find the contrasts to significantly predict conceptual change scores ( $t > 1$ ). Analyzing those participants ( $n = 96$ ; 28.57 %) who had held both misconceptions in the pre-test (agreeing to at least one misconception item per scale with high certainty; see also Dersch et al., 2022b) and hence received the feedback of having these misconceptions before reading the refutation texts, results revealed that

contrast #1 significantly predicted conceptual change scores,  $b = .26$ ,  $SE = .13$ ,  $t(92) = 4.16$ ,  $p = .044$ ,  $R^2 = .05$ , whereas both residual contrasts did not,  $t < 1$ . This finding supports the Personalized Refutation Text Hypothesis assuming that personalized refutation text most prominently fosters teachers' conceptual change for participants with misconceptions. For the result pattern see Figure 2.

**Behavioral Measure.** To test if conceptual change influenced the behavioral measure of awareness towards math-gender stereotype-reinforcing behaviors, we analyzed all participants who provided enough data in the behavioral measure to be included ( $n = 266$ ) applying the same contrasts of personalized refutation texts being superior in comparison to personalized expository text, common refutation text and common expository text (contrast #1: [-1 -1 -1 +3]) as well as comparing personalized expository text to common refutation text and common expository text (contrast #2: [-1 -1 +2 0]) and comparing common refutation text and common expository text (contrast #3: [-1 +1 0 0]) with the dependent variable of awareness towards math-gender stereotype-reinforcing behaviors. Neither contrast predicted awareness,  $ts < 1$ . These results do not support the Personalized Refutation Text Hypothesis for the behavioral measure.

We further analyzed all participants, who had held both misconceptions in the pre-test of those who provided data for the behavioral measure ( $n = 82$ ; 30.8 %). Results revealed that the first contrast testing personalized refutation texts against the other conditions significantly predicted lower awareness among those participants with both misconceptions in the pre-test,  $b = -.09$ ,  $SE = .04$ ,  $t(78) = -2.01$ ,  $p = .047$ ,  $R^2 = .06$ , suggesting backfire effects in the behavioral measure for this subsample ( $n = 82$ ). The combination of the second and third contrast testing refutation texts against common expository texts and personalized expository text against both non-personalized text forms did not predict effects in the behavioral measure,  $t < 1$ .

These results suggest that, among those participants with both misconceptions in the pre-test, personalized refutation texts may have caused backfire effects regarding awareness towards math-gender stereotype-reinforcing behaviors.

## ***H2. Shame-drives-change Hypothesis***

**Conceptual Change.** We tested whether personalized refutation text would trigger negative affect related to internal blame, which would then drive conceptual change. Therefore, we applied a mediation analysis with the contrast for personalized refutation text vs. the other conditions as independent variable (contrast: [-1 -1 -1 +3]), reported feelings of shame and guilt as a mediator, and conceptual change scores as dependent variable. Results did not reveal a significant mediation for all 336 participants, indirect effect:  $ab = -.003$ ,  $SE = .01$ , BCa CI 95 [-.02, .01],  $R^2 = .001$ , nor specifically for participants who had held both misconceptions in the pre-test ( $n = 96$ , 28.57 %), indirect effect:  $ab = -0.2$ ,  $SE = .02$ , BCa CI 95 [-.07, .02],  $R^2 = .01$ . However, we did find a small but

significant positive correlation of internal blame and conceptual change scores analyzing all 336 participants,  $r = .14, p = .009$ . This suggests that while stronger internal blame was related to more conceptual change, this internal blame was not increased by personalized refutation texts, not supporting the Shame-drives-change Hypothesis.

**Behavioral Measure.** We had assumed that more awareness towards math-gender stereotype-reinforcing behaviors in the personalized refutation text condition might be mediated by more negative affect related to internal blame. Since we did not find any positive condition effects for this behavioral measure, we did not test this hypothesis.

### *H3. Knowledge Backfire Hypothesis*

**Conceptual Change.** To test for knowledge backfire effects, meaning math-gender misconceptions being reinforced after or due to the intervention, we conducted a 2x2-ANOVA. The dependent variable was (negative) conceptual change scores with personalization vs. non-personalization and refutation vs. expository as factors, to determine which conditions were most likely to trigger the strongest knowledge backfire effects. We applied an ANOVA, because we did not know how exactly and in which condition(s) backfire effects would occur.

Results revealed a significant dysordinal interaction effect for personalization with refutation,  $F(1, 38) = 8.11, p = .007, \eta_p^2 = .18$ , while there were no main effects for personalization nor refutation ( $ps > .05$ ). These findings suggest that personalized refutation texts caused the strongest knowledge backfire effects among those participants who backfired, while common refutation text and personalized expository text did not. Overall, 42 participants' (12.5 %) conceptual change was negative, meaning their post-test math-gender misconceptions were higher than their pre-test math-gender misconceptions, that is knowledge backfire effects. However, only eight of the participants who backfired held both misconceptions in the pre-test, 34 (10.1 %) either held only one misconception or no misconception at all. Excluding those eight participants, who held both misconceptions, we still found a significant interaction effect for personalized refutation texts causing the most backfire effects within the remaining participants ( $n = 34; 10.1 \%$ ),  $F(30) = 6.05, p = .020, \eta_p^2 = .17$ . For the result pattern, see Figure 3. Our findings indicate knowledge backfire effects in the personalized refutation text condition, as assumed in the Knowledge Backfire Hypothesis.

**Behavioral Measure.** We found awareness towards math-gender stereotype-reinforcing behaviors for participants in the personalized refutation text condition with both misconceptions in the pre-test ( $n = 82; 30.8 \%$ ) to be significantly worse than in the remaining conditions,  $b = -.09, SE = .04, t(78) = -2.01, p = .047, R^2 = .06$  (see also Personalized Refutation Text Hypothesis, Behavioral Measure). This suggests backfire effects for awareness, supporting the Knowledge Backfire Hypothesis for the behavioral measure.

#### ***H4. Anger-mediates-knowledge-backfire Hypothesis***

**Conceptual Change.** We expected knowledge backfire to be mediated by negative affect related to external blame. This would be expressed by participants who read (personalized) refutation texts experiencing more negative affect related to external blame (e.g., anger). This increased affect of external blame would lead to less conceptual change among these participants in comparison to those who learn from (non-personalized) expository texts. We conducted a mediation analyses with those participants who backfired ( $n = 42$ ; 12.5 %). While our analyses revealed a direct effect,  $c' = -.24$ ,  $SE = .08$ ,  $p = .006$ , it did however not reveal an indirect effect,  $ab = -.0002$ ,  $SE = .01$ , BCa CI 95 [-.01, .03], indicating no mediation effect. Additionally, among those participants who backfired in conceptual change ( $n = 42$ ), negative affect of external blame did not differ significantly between condition,  $F_s < 1$ . For those participants who backfired, negative affect of external blame was furthermore unrelated to conceptual change,  $r = -.061$ ,  $p = .701$ . Anger did hence not mediate knowledge backfire, meaning it did not drive negative conceptual change.

**Behavioral Measure.** Since we found backfire effects for awareness towards math-gender stereotype-reinforcing behaviors among those participants who had held both misconceptions in the pre-test ( $n = 82$ ; 30.8 %) we applied a mediation analysis to test if this backfire effect was driven by negative affect of external blame. Results did not reveal a significant mediation, indirect effect:  $ab = -.003$ ,  $SE = .01$ , BCa CI 95 [-.02, .01],  $R^2 = .07$ . Furthermore, negative affect of external blame was not related to awareness,  $p > .05$ . These results do not support the Anger-mediates-knowledge-backfire Hypothesis for the behavioral measure of awareness towards math-gender stereotype-reinforcing behaviors.

#### **Preregistered Peripheral Analyses**

##### ***H5. Conceptual-change-makes-attitude-change Hypothesis***

We explored if conceptual change on math-gender misconceptions would reduce the math-gender stereotype or foster positive attitudes regarding closing the math-gender gap. We applied the same contrasts as before (#1: [-1 -1 -1 +3] for our main explanation, #2: [-1 -1 +2 0] and #3: [-1 +1 0 0] both to rule out alternative explanations) with the dependent variables of the global math-gender stereotype and specific positive attitudes towards gender equity in mathematics. We tested whether the most effective condition for fostering conceptual change—personalized refutation text among those participants who had held both misconceptions in the pre-test ( $n = 96$ ; 28.6%)—would reduce the math-gender stereotype or foster specific positive attitudes towards gender equity in mathematics compared to other conditions. Analyzing those participants who had held both misconceptions in the pre-test as well as all 336 participants, we found that personalized refutation texts did not foster either attitude change more than the other conditions,  $ps > .05$ . Furthermore, since we did not measure attitudes before and after

the intervention, we report correlations here. For all participants, math-gender misconceptions at the post-test were negatively and significantly correlated with positive attitudes towards gender equity in mathematics,  $r = -.20$ ,  $p < .001$ , implying that lower levels of math-gender misconceptions at post-test, predicted more positive attitude towards equity in mathematics and closing the math gender gap. These findings point toward the direction of the Conceptual-change-makes-attitude-change Hypothesis.

#### ***H6. Attitude Backfire Hypothesis***

We explored whether participants who read the (personalized) refutation texts had a more negative attitude towards the math-gender gap being closed or stronger a math-gender stereotype than participants who read the (non-personalized) expository texts about math-gender misconceptions. Namely, we conducted a 2x2-MANOVA with all 336 participants, with the two attitude measures of the math-gender stereotype as well as participants' attitude towards (closing) the math-gender gap as dependent variables and personalization vs. non-personalization and refutation vs. expository as factors. We found no main effects nor an interaction effect ( $ps < .05$ ). We further conducted the same MANOVA with participants who had held both misconceptions in the pre-test ( $n = 96$ ; 28.6 %). Here, we did not find any effects either ( $ps < .05$ ). However, analyzing all participants with knowledge backfire ( $n = 42$ ; 12.5 %), we found a main effect for personalization on specific positive attitudes towards gender equity in mathematics,  $F(1, 32) = 5.92$ ,  $p = .021$ ,  $\eta_p^2 = .16$ , suggesting that attitude backfire was weakest and hence attitudes towards closing the math-gender gap most positive for those participants in both personalized conditions, personalized expository text and personalized refutation text. However, while descriptively going to a similar direction, this effect was not significant for the general math-gender stereotype ( $p > .05$ ) and hence missed significance within the MANOVA with both dependent variables ( $p = .05$ ). These findings suggest that personalization may have a positive effect for those participants with knowledge backfire to prevent attitude backfire or even improve attitude in comparison to the other conditions. However, this holds true only for more specific attitudes towards the math-gender gap, not for the general math-gender stereotype.

#### ***H7. Anger-mediates-attitude-backfire Hypothesis.***

We hypothesized that attitude backfire would be mediated by negative affect related to external blame (e.g., anger). Since we did not find evidence for attitude backfire, we did not conduct this mediation analysis.

### **Discussion**

We examined whether teachers' conceptual change is strongest using personalized refutation text in a 2x2 design comparing it to non-personalized refutation text, personalized expository text and non-personalized expository text. We were interested if teachers' math-gender misconceptions could be successfully reduced using personalized refutation text or if personalized expository text or common refutation would be sufficient for teachers' conceptual

change. Teachers being a challenging target group for conceptual change interventions (Gregoire, 2003) and math-gender misconceptions being an attitude-prone (e.g., Abrams et al., 1990), social-scientific topic likely triggering worldview backfire (Jacobson et al., 2021), we also accounted for backfire effects (e.g., Lewandowsky, 2012). Furthermore, we aimed to understand the role of affect related to internal blame and external blame for successful conceptual change and for backfire effects. Lastly, we explored the potential effects of our intervention(s) on attitude and potential attitude backfire.

### **Personalized Refutation Texts Can Foster Conceptual Change**

We found that participants benefited most from personalized refutation text in comparison to common refutation text, personalized expository text or common expository text. However, we only found this effect for participants who had held both misconceptions in the pre-test. This means, for these participants, personalized refutation texts that combined personalized feedback about their misconception with a refutation text targeting this misconception, fostered the most conceptual change. This finding is in line with theoretical approaches on teachers' conceptual change, which suggest that teachers need additional self-implication for successful conceptual change (e.g., Gregoire, 2003). Additionally, our finding is in line with previous research suggesting that personalized refutation instruction can be a powerful tool revising misconceptions (Asterhan & Dotan, 2018), especially for teachers (Dersch et al., 2022b).

Findings also suggest that, for teachers who hold misconceptions, neither personalized expository texts nor common refutation texts are enough to foster their conceptual change more than common, non-personalized expository text. In other words, for this specific group, personalized refutation texts can be confirmed to best foster conceptual change (see also Dersch et al., 2022b).

### **Backfire Effects**

However, the positive effect of personalized refutation texts was not present among *all* participating teachers, but only among those *with* misconceptions. This finding may be due to backfire effects, which we found to be strongest in the personalized refutation text condition. Yet, contrary to what we had assumed, worldview backfire was most likely not the kind of backfire that occurred, since we found backfire effects to be strongest for participants who had held no or only few misconceptions at pre-test. Contrary to our expectations, *worldview* backfire is unlikely to have occurred because it likely occurs among people who hold misconceptions and are certain in their misconceptions (Eitel et al., 2021; Lewandowsky et al., 2020). However, participants who backfired were uncertain or did not hold any misconceptions in the pre-test. Worldview backfire would furthermore likely be related to negative affect related to external blame (Trevors et al., 2016). However, participants who backfired did not hold strong misconceptions and their backfire effects were not related to negative affect of external blame.

Hence, we likely found *familiarity backfire*. Familiarity backfire occurs when the mere mentioning or describing of a misconception reinforces the misconception by making it more familiar to people (Lewandowsky et al., 2020). Evidence on familiarity backfire is mixed, while newer research could not replicate it (Ecker et al., 2023). What might however still have led to familiarity backfire in this study, especially among the subgroup of participants who received personalized refutation texts while not holding misconceptions prior, is the combination of (1) not being familiar with the misconception before being confronted with it (see also Ecker et al., 2023), (2) being confronted with the misconception not once, as in common refutation texts (see Kendoue & O'Brien, 2014), but twice. Namely, participants in the personalized refutation condition first received feedback about not holding a misconception, before reading the refutation text which mentioned the misconception again. Therefore, in this specific case the increased familiarity was greater than the refutational benefit, which rarely happens, but still likely causes familiarity backfire (Lewandowsky et al., 2020).

We further found signs of backfire effects in the behavioral measure of awareness towards teacher behaviors that might reinforce the math-gender stereotype in the classroom. Here, personalized refutation text led to less awareness towards math-gender stereotype-reinforcing behaviors than common expository text, refutation text or personalized expository text. Our findings here are similar to prior findings, which did not reveal effects of (personalized) refutation text on behavioral measures among student teachers (Prinz et al., 2022) and in-service teachers (Dersch et al., 2022b; Ferrero et al., 2020a). In one study, Ferrero et al. (2020b), similar to our findings, also found backfire effects in a behavioral measure. One possible explanation is the *continued influence effect* (Seifert, 2014). This effect describes individuals oftentimes relying on misconceptions in their reasoning and decision making—for instance in a behavioral measure like here—even after a successful refutation of said misconception (Seifert, 2014).

### **The Role of Negative Affect of Internal or External Blame**

In this sample, we found neither negative affect of internal blame, such as guilt or shame, nor negative affect of external blame, such as anger, to mediate teachers' conceptual change or backfire affects when targeting teachers' math-gender misconceptions. For external blame, our findings are in line with previous findings, who did also not find negative affect of external blame to influence findings (Dersch et al., 2022b; Prinz et al., 2022). For internal blame, our findings are, however, not in line with previous research, which, for the topic of multimedia learning did find internal blame to mediate conceptual change (Dersch et al., 2022b). While we did find a relation of negative affect of internal blame with conceptual change, this was seemingly not influenced by the condition participants were in. It may have rather been due to the topic itself, because related attitudes are likely stronger for the social-scientific topic of math-gender misconceptions (e.g., Abrams et al., 1990; Crain et al., 1996; Jacobson et

al., 2021). Hence, not the manipulation, but the topic triggered negative affect of internal blame, which then may have enhanced conceptual change. Personalized refutation texts were, however, the most effective condition at fostering conceptual change. This might be due to other processes than the attempted increase in self-implication. Personalization may have led to the activation of self-schemata improving the integration of the to-be-learned information in these prior knowledge structures (Markus, 1977). Furthermore, meta-cognition (Vosniadou, 1994) and cognitive co-activation of the misconceptions and the alternative explanation (Kendeo & O'Brien, 2014), may have been enhanced by personalized feedback being combined with refutation text.

### **Potential Effects on Attitude**

We also explored if conceptual change of math-gender misconceptions might reduce the math-gender stereotype or lead to more positive attitudes towards closing the math-gender gap. We did not find condition effects. We did however find a negative relation between math-gender misconceptions at post-test and the more specific, equity-oriented attitude towards the math-gender gap. Namely, we found that lower levels of misconceptions at post-test were related to more positive attitudes towards equity in mathematics and closing the math gender gap. We did not find relationships between misconception levels and the global math-gender stereotype. Our findings are in line with previous evidence that teachers' attitudes about gender disparities in society are associated with unsupported gender-specific ability beliefs (Copur-Gencturk et al., 2023). Further, our findings are consistent with research which found that the math-gender stereotype cannot be changed easily and sustainably (FitzGerald et al., 2019; Kollmayer et al., 2020). To change the general math-gender stereotype, we would likely need a more extensive intervention, similar to Kollmayer et al. (2020), or an intervention with multiple treatments over time. We however found that personalization, among participants who experienced *knowledge* backfire, led to an overall positive attitude towards closing the math-gender gap and gender equity in STEM subjects. Hence, personalization might be a protective factor from attitude backfire.

### **Limitations and Further Research**

This study being an online study holds some disadvantages. We did not find a lot of variance in the negative affect of external blame among our participants ( $M = 1.23$  [on a scale of 1 to 5];  $SD = 0.45$ ). While one possible explanation might be the topic simply does not cause any negative affect of external blame, which might contradict the literature to some extent (see Abrams et al., 1990; Arens et al., 2021; Crain et al., 1996; Jacobson et al., 2021; Trevors et al., 2016), it is also possible that the lack of variance in negative affect related to external blame may be due to dropout. Some participants due to experiencing negative affect of external blame during the study may have simply left the study. However, this seems to be the case regardless of the condition the participants were in, since dropout was equally distributed among conditions. This hints towards the topic itself being the reason for dropout

rather than the kind of manipulation, which may have been too weak for people with prior strong attitudes regarding mathematics and gender. Research on conceptual change about math-gender misconceptions in the laboratory is thus necessary. In the laboratory, it would further be of interest to also measure peripheral physiological data like heart rate variability or facial micro emotions (e.g., Pourmohammadi & Maleki, 2020), which might be interesting additional data regarding perceived stress when confronted with own misconceptions.

However, this study being an online study is also a strength because in-service teachers who have limited time-resources (Creagh et al., 2023), are unlikely to come to a university lab to take part in a study. Due to the study being online, we were able to collect a sample which also contained in-service teachers. Furthermore, online interventions might be an economic way to provide teachers, who have little time resources, with ‘snackable’ professional development opportunities, conceptual change of potential misconceptions being an interesting focal point of such professional development online interventions.

The measure of awareness towards math-gender stereotype-reinforcing behaviors in a model teacher in previous research had been weakly but significantly negatively correlated with math-gender misconceptions (Authors, submitted) making it an appropriate transfer measure. However, here, we could not replicate this correlation, which is why we treat it as an additional behavioral measure instead of a transfer measure. Hence, in further research, this behavioral measure should be improved, for instance by scripting and filming videos that are closer to the content of the math-gender misconceptions. Nevertheless, we think that the current behavioral measure adds value. Especially, since conceptual change does not exist in a vacuum, but likely influences behaviors (see also Dersch et al., 2022b; Ferrero et al., 2020a; 2020b; Prinz et al., 2022). Researching potentially related behaviors is especially important in teachers as multipliers and important socializers with an impact on their students’ (math-gender) stereotypes (Eccles, 2011; Gunderson et al., 2012).

There are two noteworthy limitations regarding our sample. First, the sample did not contain only mathematics teachers, but all kinds of teachers. Of all 336 participants, 213 indicated to teach mathematics or study to teach mathematics. That is still a substantial part of our sample. To make sure that mathematics as a subject did not influence misconceptions in the pre-test, we compared mathematics teachers with teachers of all remaining subjects. Mathematics teachers did not differ from other teachers regarding their math-gender misconceptions in the pre-test ( $p > .05$ ). Additionally, for teachers who do not teach mathematics, it may still be relevant to reduce their misconceptions about mathematics and gender, since all teachers are important socializers and may still influence their students’ math-gender stereotype (Gunderson et al., 2012).

The second limitation of our sample is that men were underrepresented in comparison to the general population. Only 20.5 % of all participants were socialized as male. However, this is quite representative of teachers

in Germany, who, according to the German Federal Office for Statistics in the school year of 2019/2020 were 73.1 % female. Math-gender misconceptions in the pre-test furthermore did not differ between the genders ( $p > .05$ ).

### **Conclusion and Implications for Theory and Practice**

Our findings suggest that personalized refutation text among the interventions we implemented is the most potent strategy for refuting teachers' math-gender misconceptions. This effect seems to be consistent across different topics (Dersch et al., 2022b). However, personalized refutation text is only effective if misconceptions are prevalent. If they are not prevalent, personalized refutation text leads to familiarity backfire effects due to mentioning the misconception several times in this intervention without the need for refutation (Lewandowsky et al., 2020). Personalized feedback in a refutation intervention should hence be combined with an adaptive algorithm, which only provides a refutation text if a misconception is prevalent and, in other cases, provides an expository text, which has been done by Dersch et al. (2022b).

We found that personalization, among those participants who experienced knowledge backfire, led to the most positive attitude towards closing the math-gender gap in comparison to the non-personalized conditions. This suggests that personalization may be a protective factor from attitude backfire effects. While our results suggest a general positive effect of all kinds of intervention on specific attitudes about closing the math-gender gap, for a reduction of the global math-gender stereotype, more extensive interventions are necessary (Kollmayer et al., 2020).

We found signs of backfire effects in our behavioral measure of awareness towards math-gender stereotype-reinforcing behaviors in the classroom for personalized refutation text, which, for participants with misconceptions, led to less awareness in comparison to the other conditions, which is consistent with prior findings by Ferrero et al. (2020b). Although personalized refutation texts might trigger behavioral backfire, they also have the potential to reduce math-gender misconceptions. Intervention designers should thus carefully consider if the focus of their efforts is to reduce misconceptions or if potentially related behaviors and attitudes are also important.

To answer the questions raised by the title of this paper: Should we tell teachers that they are correct? In this case, should we tell teachers who do not hold misconceptions about mathematics and gender that they hold no misconceptions? Yes, but only if we also refrain from providing a refutation text afterwards. If so, then personalization can still be beneficial for positive attitude changes. Should we tell teachers that they are incorrect? Yes, if they hold misconceptions, they should know; and additionally, we should combine that feedback with a refutation text. However, we should take the possibility of behavioral backfire into account.

### References

- Abrams, D., Thomas, J., & Hogg, M. A. (1990). Numerical distinctiveness, social identity and gender salience. *British Journal of Social Psychology*, 29(1), 87-92. <https://doi.org/10.1111/j.2044-8309.1990.tb00889.x>
- Altman, D. G. (1990). *Practical statistics for medical research*. Chapman and Hall/CRC.
- Arens, A. K., Jansen, M., Preckel, F., Schmidt, I., & Brunner, M. (2021). The structure of academic self-concept: A methodological review and empirical illustration of central models. *Review of educational research*, 91(1), 34-72. <https://doi.org/10.3102/0034654320972186>
- Ashmore, R. D., & Del Boca, F. K. (1979). Sex stereotypes and implicit personality theory: Toward a cognitive—Social psychological conceptualization. *Sex roles*, 5(2), 219-248. <https://doi.org/10.1007/BF00287932>
- Asterhan, C. S., & Dotan, A. (2018). Feedback that corrects and contrasts students' erroneous solutions with expert ones improves expository instruction for conceptual change. *Instructional Science*, 1-19. <https://doi.org/10.1007/s11251-017-9441-1>
- Baron-Cohen, S. (2005, January). The essential difference: The male and female brain. *Phi Kappa Phi Forum*, 85(1), 23—26.
- Bodenhausen, G. V. (1990). Stereotypes as Judgmental Heuristics: Evidence of Circadian Variations in Discrimination. *Psychological Science*, 1(5), 319-322. <https://doi.org/10.1111/j.1467-9280.1990.tb00226.x>
- Brookover, W. B., Thomas, S., & Paterson, A. (1964). Self-concept of ability and school achievement. *Sociology of education*, 271-278. <https://doi.org/10.2307/2111958>
- Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics*, 134(3), 1163-1224. <https://doi.org/10.1093/qje/qjz008>
- Chi, M. T. (2009). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In *International handbook of research on conceptual change* (pp. 89-110). Routledge.
- Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of educational research*, 63(1), 1-49. <https://doi.org/10.3102/00346543063001001>
- Cook, J., & Lewandowsky, S. (2011). *The Debunking Handbook*. University of Queensland. <http://sks.to/debunk>

- Copur-Gencturk, Y., Thacker, I., & Cimpian, J. R. (2023). Teachers' race and gender biases and the moderating effects of their beliefs and dispositions. *International Journal of STEM Education*, 10(1), 31.  
<https://doi.org/10.1186/s40594-023-00420-z>
- Crain, R. M. (1996). The influence of age, race, and gender on child and adolescent multidimensional self-concept. <https://doi.org/10.1080/10573560308222>
- Creagh, S., Thompson, G., Mockler, N., Stacey, M., & Hogan, A. (2023). Workload, work intensification and time poverty for teachers and school leaders: A systematic research synthesis. *Educational Review*, 1-20. <https://doi.org/10.1080/00131911.2023.2196607>
- Danielson, R. W., Jacobson, N. G., Patall, E. A., Sinatra, G. M., Adesope, O. O., Kennedy, A. A., ... & Sunday, O. J. (2024). The effectiveness of refutation text in confronting scientific misconceptions: A meta-analysis. *Educational Psychologist*, 1-25. <https://doi.org/10.1080/00461520.2024.2365628>
- Dersch, A.-S., Heyder, A., & Eitel, A. (2022a). Exploring the Nature of Teachers' Math-gender Stereotypes: The Math-gender Misconception Questionnaire. *Frontiers in psychology*, 1728.  
<https://doi.org/10.3389/fpsyg.2022.820254>
- Dersch, A.-S., Renkl, A., & Eitel, A. (2022b). Personalized refutation texts best stimulate teachers' conceptual change about multimedia learning. *Journal of Computer Assisted Learning*.  
<https://doi.org/10.1111/jcal.12671>
- DiSessa, A. A. (1988). Knowledge in pieces.
- Dole, J. A., & Sinatra, G. M. (1998). Reconceptualizing change in the cognitive construction of knowledge. *Educational psychologist*, 33(2-3), 109-128.  
<https://doi.org/10.1080/00461520.1998.9653294>
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt brace Jovanovich college publishers.
- Eccles, J. (2011). Gendered educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *International Journal of Behavioral Development*, 35(3), 195-201.  
<https://doi.org/10.1177%2F0165025411398185>
- Eccles, J., Adler, T. F., Futtermann, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). "Expectations, values and academic behaviors," in *Perspective on achievement and achievement motivation*, ed. J. T. Spence (Cambridge, MA: Academic Press), 75-146.
- Ecker, U. K., Swire, B., & Lewandowsky, S. (2014). Correcting misinformation—A challenge for education and cognitive science. <https://doi.org/10.7551/mitpress/9737.003.0005>

- Ecker, U. K., Sharkey, C. X., & Swire-Thompson, B. (2023). Correcting vaccine misinformation: A failure to replicate familiarity or fear-driven backfire effects. *Plos one*, *18*(4), e0281140.  
<https://doi.org/10.1371/journal.pone.0281140>
- Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of personality and social psychology*, *71*(1), 5.
- Eitel, A., Dersch, A.-S., & Renkl, A. (2019). Wissenschaftliche Annahmen und Fehlkonzepte angehender Lehrkräfte über das Lernen mit Multimedia. *Unterrichtswissenschaft*, *1*-24.  
<https://doi.org/10.1007/s42010-019-00049-4>
- Eitel, A., Prinz, A., Kollmer, A., Niessen, L., Russow, J., Ludäscher, M., Renkl, A. & Lindner, M. (2020). *The Misconceptions about Multimedia Learning Questionnaire (MMLQ): Development and Validation with Teachers and Student Teachers*. Manuskript submitted for publication.
- Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*(1), 103-127.  
<https://doi.org/10.1037/a0018053>
- Escovar, E., Rosenberg-Lee, M., Uddin, L. Q., & Menon, V. (2016). The empathizing-systemizing theory, social abilities, and mathematical achievement in children. *Scientific reports*, *6*(1), 1—11.  
<https://doi.org/10.1038/srep23011>
- Ferrero, M., Konstantinidis, E., & Vadillo, M. A. (2020a). An attempt to correct erroneous ideas among teacher education students: The effectiveness of refutation texts. *Frontiers in psychology*, *11*, 577738.  
<https://doi.org/10.3389/fpsyg.2020.577738>
- Ferrero, M., Hardwicke, T. E., Konstantinidis, E., & Vadillo, M. A. (2020b). The effectiveness of refutation texts to correct misconceptions among educators. *Journal of experimental psychology: applied*, *26*(3), 411.  
<https://psycnet.apa.org/doi/10.1037/xap0000258>
- Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Stanford University Press.  
<https://doi.org/10.1038/scientificamerican1062-93>
- Fiske, S. T. (1982). Schema-triggered affect: Applications to social perception. In *Affect and cognition: 17th Annual Carnegie Mellon symposium on cognition* (pp. 55-78). Hillsdale: Lawrence Erlbaum.
- FitzGerald C., Martin A., Berner D., Hurst S. (2019). Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: a systematic review. *BMC Psychol.* *7*, 1–12.  
<https://doi.org/10.1186/s40359-019-0299-7>

- Gini, S., Knowland, V., Thomas, M. S., & Van Herwegen, J. (2021). Neuromyths about neurodevelopmental disorders: Misconceptions by educators and the general public. *Mind, Brain, and Education, 15*(4), 289-298.
- Girelli, L. (2022). What does gender has to do with math? Complex questions require complex answers. *Journal of Neuroscience Research*. <https://doi.org/10.1002/jnr.25056>
- Gottfredson, L. S. (1998). The general intelligence factor.
- Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological review, 109*(1), 3. <https://doi.org/10.1037/0033-295X.109.1.3>
- Gregoire, M. (2003). Is it a challenge or a threat? A dual-process model of teachers' cognition and appraisal processes during conceptual change. *Educational psychology review, 15*, 147-179. <https://doi.org/10.1023/A:1023477131081>
- Grospietsch, F., & Mayer, J. (2018). Professionalizing pre-service biology teachers' misconceptions about learning and the brain through conceptual change. *Education Sciences, 8*(3), 120. <https://doi.org/10.3390/educsci8030120>
- Gunderson, E. A., Ramirez, G., Levine, S.C, & Beilock, S. L., (2012). The Role of Parents and Teachers in the Development of Gender-Related Math Attitudes. *Sex Roles, 66*, 153–166 (2012). <https://doi.org/10.1007/s11199-011-9996-2>
- Hand, S., Rice, L., & Greenlee, E. (2017). Exploring teachers' and students' gender role bias and students' confidence in STEM fields. *Social Psychology of Education, 20*(4), 929-945. <https://doi.org/10.1007/s11218-017-9408-8>
- Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychol. Sci. Public Inter. 8*, 1–51. <https://10.1111/j.1529-1006.2007.00032.x>
- Hughes, S., Lyddy, F., and Lambe, S. (2013). Misconceptions about Psychological Science: A Review. *Psychology Learning & Teaching, 12*, 20 - 31. <https://doi.org/10.2304/plat.2013.12.1.20>
- Jacobson, N. G., Thacker, I., & Sinatra, G. M. (2021). Here's hoping it's not just text structure: The role of emotions in knowledge revision and the backfire effect. *Discourse Processes, 59*(1-2), 13-35. <https://doi.org/10.1080/0163853X.2021.1925059>
- Kendeou, P. (2024). A Theory of Knowledge Revision: the Development of the KReC Framework. *Educational Psychology Review, 36*(2), 44. <https://doi.org/10.1007/s10648-024-09885-y>

- Kendeou, P., Smith, E. R., & O'Brien, E. J. (2013). Updating during reading comprehension: why causality matters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 854
- Kendeou, P., Walsh, E. K., Smith, E. R., & O'Brien, E. J. (2014). Knowledge revision processes in refutation texts. *Discourse Processes*, *51*(5-6), 374-397. <https://doi.org/10.1080/0163853X.2014.913961>
- Kendeou, P., & O'Brien, E. J. (2014). The knowledge revision components (KReC) framework: Processes and mechanisms. <https://doi.org/10.7551/mitpress/9737.003.0022>
- Kersey, A. J., Csumitta, K. D., & Cantlon, J. F. (2019). Gender similarities in the brain during mathematics development. *npj Science of Learning*, *4*(1), 1—7. <https://doi.org/10.1038/s41539-019-0057-x>
- Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in psychology*, *8*, 1997. <https://doi.org/10.3389/fpsyg.2017.01997>
- Klepsch, M., & Seufert, T. (2021, April). Making an effort versus experiencing load. *Frontiers in Education*, *6*. <https://doi.org/10.3389/educ.2021.645284>
- Kollmayer M., Schultes M. T., Lüftenegger M., Finsterwald M., Spiel C., Schober B. (2020). REFLECT-a teacher training program to promote gender equality in schools. *Front. Educ.* *5*, 136. <https://doi.org/10.3389/educ.2020.00136>
- Lahance, J. A., & Mazzocco, M. M. (2006). A longitudinal analysis of sex differences in math and spacial skills in primary school age children. *Learning and individual differences*, *16*(3), 195—2016. <https://doi.org/10.1016/j.lindif.2014.03.020>
- Larkin D. (2012). Misconceptions about “misconceptions”: Preservice secondary science teachers’ views on the value and role of student ideas. *Sci. Educ.* *96*, 927–959. <https://doi.org/10.1002/sce.21022>
- Lefcourt, H. M. (1991). Locus of control. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 413–499). Academic Press. <https://doi.org/10.1016/B978-0-12-590241-0.50013-7>
- Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E. Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., Vraga, E. K., Wood, T. J., Zaragoza, M. S. (2020). The Debunking Handbook 2020. <https://doi.org/10.17910/b7.1182>
- Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological science in the public interest*, *13*(3), 106-131. <https://doi.org/10.1177/1529100612451018>

- Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychological bulletin*, 136(6), 1123. <https://doi.org/10.1037/a0021276>
- Maheswaran, D., & Chaiken, S. (1991). Promoting systematic processing in low-motivation settings: Effect of incongruent information on processing and judgment. *Journal of personality and social psychology*, 61(1), 13. <https://doi.org/10.1037/0022-3514.61.1.13>
- Maier, U., & Klotz, C. (2022). Personalized feedback in digital learning environments: Classification framework and literature review. *Computers and Education: Artificial Intelligence*, 3, 100080. <https://doi.org/10.1016/j.caeai.2022.100080>
- Markus, H. (1977). Self-schemata and processing information about the self. *Journal of personality and social psychology*, 35(2), 63. <https://psycnet.apa.org/doi/10.1037/0022-3514.35.2.63>
- Martingano, A. J., & Konrath, S. (2022). How cognitive and emotional empathy relate to rational thinking: empirical evidence and meta-analysis. *The Journal of Social Psychology*, 162(1), 143–160. <https://doi.org/10.1080/00224545.2021.1985415>
- Mason, L. (2010). Beliefs about knowledge and revision of knowledge: On the importance of epistemic beliefs for intentional conceptual change in elementary and middle school students. *Personal epistemology in the classroom: Theory, research, and implications for practice*, 258-291. <https://doi.org/10.1017/CBO9780511691904.009>
- Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2. Ed. pp. 43–71). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.005>
- McAfee, M., & Hoffman, B. (2021). The morass of misconceptions: How unjustified beliefs influence pedagogy and learning. *International Journal for the Scholarship of Teaching and Learning*, 15(1), 4. <https://doi.org/10.20429/ijstol.2021.150104>
- Moosbrugger, H., & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion*. [https://doi.org/10.1007/978-3-642-20072-4\\_1](https://doi.org/10.1007/978-3-642-20072-4_1)
- Muis, K. R., Sinatra, G. M., Pekrun, R., Winne, P. H., Trevors, G., Losenno, K. M., & Munzar, B. (2018). Main and moderator effects of refutation on task value, epistemic emotions, and learning strategies during conceptual change☆. *Contemporary Educational Psychology*, 55, 155-165. <https://doi.org/10.1016/j.cedpsych.2018.10.001>

- Nabi, R. L. (1999). A cognitive-functional model for the effects of discrete negative emotions on information processing, attitude change, and recall. *Communication theory*, 9(3), 292-320.  
<https://doi.org/10.1111/j.1468-2885.1999.tb00172.x>
- Narciss, S., Sosnovsky, S., Schnaubert, L., Andrès, E., Eichelmann, A., Gogvadze, G., & Melis, E. (2014). Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Computers & Education*, 71, 56-76. <https://doi.org/10.1016/j.compedu.2013.09.011>
- Nosek, B. A., & Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *American Educational Research Journal*, 48(5), 1125-1156.  
<https://doi.org/10.3102%2F0002831211410683>
- Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Banaji, M. R., et al. (2010). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, 18(1), 36-88. <https://doi.org/10.1080/10463280701489053>
- Ommundsen, Y., Haugen, R., & Lund, T. (2005). Academic self-concept, implicit theories of ability, and self-regulation strategies. *Scandinavian Journal of Educational Research*, 49(5), 461-474.  
<https://doi.org/10.1080/00313830500267838>
- Pourmohammadi, S., & Maleki, A. (2020). Stress detection using ECG and EMG signals: A comprehensive study. *Computer methods and programs in biomedicine*, 193, 105482.  
<https://doi.org/10.1016/j.cmpb.2020.105482>
- Prinz, A., Kollmer, J., Flick, L., Renkl, A., & Eitel, A. (2022). Refuting student teachers' misconceptions about multimedia learning. *Instructional Science*, 50(1), 89-110. <https://doi.org/10.1007/s11251-021-09568-z>
- Renkl, A. (2015). Different roads lead to Rome: The case of principle-based cognitive skills. *Learning: Research & Practice*, 1, 79-90. <https://doi.org/10.1080/23735082.2015.994255>
- Sáinz, M., Fàbregues, S., & Solé, J. (2020). Parent and teacher depictions of gender gaps in secondary student appraisals of their academic competences. *Frontiers in psychology*, 11.  
<https://dx.doi.org/10.3389%2Ffpsyg.2020.573752>
- Sánchez, E., García-Rodicio, H., & Acuña, S. R. (2009). Are instructional explanations more effective in the context of an impasse? *Instructional Science*, 37, 537-563. <https://doi.org/10.1007/s11251-008-9074-5>
- Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, 134(2), 207-222. <https://doi.org/10.1037/0033-2909.134.2.207>

- Sherman, D. K., & Cohen, G. L. (2002). Accepting threatening information: Self-affirmation and the reduction of defensive biases. *Current Directions in Psychological Science*, *11*(4), 119–123.  
<https://doi.org/10.1111/1467-8721.00182>
- Sinatra, G. M. (2005). The “warming trend” in conceptual change research: The legacy of Paul R. Pintrich. *Educational Psychologist*, *40*(2), 107–115. [https://doi.org/10.1207/s15326985ep4002\\_5](https://doi.org/10.1207/s15326985ep4002_5)
- Steele, J. (2003). Children’s gender stereotypes about math: The role of stereotype stratification 1. *Journal of applied social psychology*, *33*(12), 2587-2606. <https://doi.org/10.1111/j.1559-1816.2003.tb02782.x>
- 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), 324-335. <https://doi.org/10.1007/s11199-010-9924-x>
- Stockard, J. (2006). Gender socialization. *Handbook of the Sociology of Gender*, 215-227.  
[https://doi.org/10.1007/0-387-36218-5\\_11](https://doi.org/10.1007/0-387-36218-5_11)
- Strack, F., & Deutsch, R. (2015). The duality of everyday life: Dual-process and dual system models in social psychology. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology, Vol. 1. Attitudes and social cognition* (pp. 891–927). American Psychological Association. <https://doi.org/10.1037/14341-028>
- Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Academic Press. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8> \t "\_blank" \o "Persistent link using digital object identifier
- Sweller, J., Van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, *10*, 251-296. <https://doi.org/10.1023/A:1022193728205>
- Tiedemann, J. (2000). Parents’ gender stereotypes and teachers’ beliefs as predictors of children’s concept of their mathematical ability in elementary school. *Journal of Educational psychology*, *92*(1), 144.  
<https://psycnet.apa.org/doi/10.1037/0022-0663.92.1.144>
- Tiedemann, J. (2002). Teachers’ gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educational Studies in mathematics*, *50*(1), 49-62.  
<https://doi.org/10.1023/A:1020518104346>
- Tippett, C. D. (2010). Refutation text in science education: A review of two decades of research. *International Journal of Science and Mathematics Education*, *8*(6), 951-970. <https://doi.org/10.1007/s10763-010-9203-x>

- Trevors, G. J., Muis, K. R., Pekrun, R., Sinatra, G. M., & Winne, P. H. (2016). Identity and epistemic emotions during knowledge revision: A potential account for the backfire effect. *Discourse Processes, 53*(5-6), 339-370. <https://doi.org/10.1080/0163853X.2015.1136507>
- Trevors, G., & Kendeou, P. (2020). The effects of positive and negative emotional text content on knowledge revision. *Quarterly Journal of Experimental Psychology, 73*(9), 1326-1339. <https://doi.org/10.1177/1747021820913816>
- von Aufschnaiter, C. & Rogge, C. (2015). Conceptual Change in Learning. In R. Gunstone (Ed.), *Encyclopedia of Science Education* (pp. 209-218). Springer.
- Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and instruction, 4*(1), 45-69. [https://doi.org/10.1016/0959-4752\(94\)90018-3](https://doi.org/10.1016/0959-4752(94)90018-3)
- Vosniadou, S., Lawson, M. J., Wyra, M., Van Deur, P., Jeffries, D., & Ngurah, D. I. G. (2020). Pre-service teachers' beliefs about learning and teaching and about the self-regulation of learning: A conceptual change perspective. *International Journal of Educational Research, 99*, 101495. <https://doi.org/10.1016/j.ijer.2019.101495>
- Vosniadou, S., Darmawan, I., Lawson, M. J., Van Deur, P., Jeffries, D., & Wyra, M. (2021). Beliefs about the self-regulation of learning predict cognitive and metacognitive strategies and academic performance in pre-service teachers. *Metacognition and Learning, 1-32*. <https://doi.org/10.1007/s11409-020-09258-0>
- Wang, M.-T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review, 29*, 119—140. <https://doi.org/10.1007/s10648-015-9355-x>
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology, 54*(6), 1063 - 1070. <https://doi.org/10.1037/0022-3514.54.6.1063>
- Weiner, B. (1985). Attribution theory. *Human motivation, 275-326*.
- Wilén, W. W. (2004). Refuting misconceptions about classroom discussion. *The Social Studies, 95*(1), 33-39. <https://doi.org/10.3200/TSSS.95.1.33-39>
- Zengilowski, A., Schuetze, B. A., Nash, B. L., & Schallert, D. L. (2021). A critical review of the refutation text literature: Methodological confounds, theoretical problems, and possible solutions. *Educational Psychologist, 56*(3), 175-195. <https://doi.org/10.1080/00461520.2020.1861948>

**Table 1**

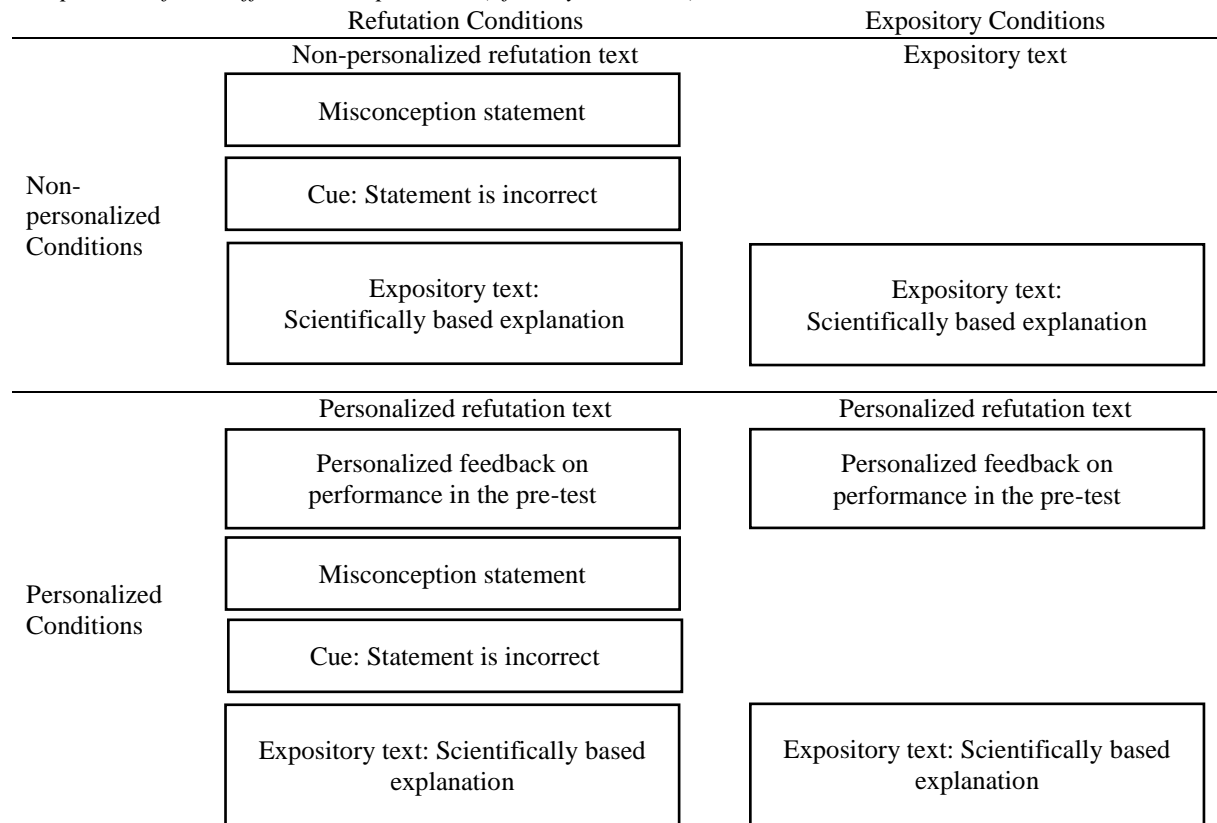
*Means (with standard deviations) and distributions for the control variables as a function of experimental condition*

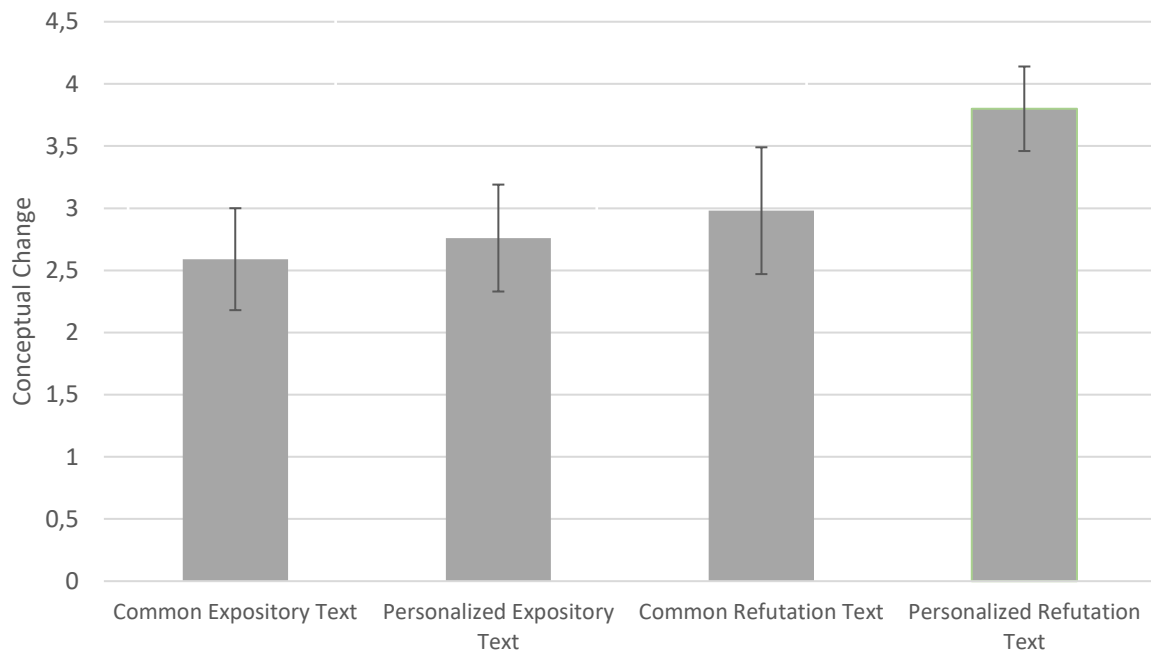
	<b>Expository Texts</b>	<b>Refutation Texts</b>	<b>Personalized Expository Texts</b>	<b>Personalized Refutation Texts</b>	<b>All</b>
Gender (female / male / non-binary)	62 / 21 / 0	61 / 13 / 0	70 / 15 / 1	73 / 20 / 0	266 / 69 / 1
German as Mother Language (yes / no)	82 / 1	67 / 7	82 / 4	88 / 5	319 / 17
Age in Years	27.76 (13.95)	25.95 (11.42)	27.50 (11.54)	28.71 (13.81)	27.56 (12.77)
Status (student teacher / student teacher who teaches / fully examined teacher)	46 / 15 / 22	50 / 10 / 14	54 / 7 / 25	51 / 18 / 24	201 / 50 / 85
Teaching Experience in Years	6.05 (10.80)	3.72 (8.21)	4.36 (7.92)	5.96 (10.54)	5.08 (9.52)
School Type (primary / secondary)	37 / 46	25 / 49	34 / 52	36 / 57	132 / 204
ICL <sup>a</sup> (min. = 1, max. = 7)	2.67 (1.26)	2.73 (1.09)	2.90 (1.08)	2.84 (1.25)	2.79 (1.17)
ECL <sup>a</sup> (min. = 1, max. = 7)	2.24 (1.09)	2.18 (1.00)	2.35 (1.09)	2.30 (1.02)	2.27 (1.05)
Pretest Scores	-0.78 (2.05)	-0.77 (1.77)	-0.53(1.84)	-0.70 (1.77)	-0.69 (1.86)

*Notes.* <sup>a</sup> ICL = Intrinsic Cognitive Load; ECL = Extraneous Cognitive Load

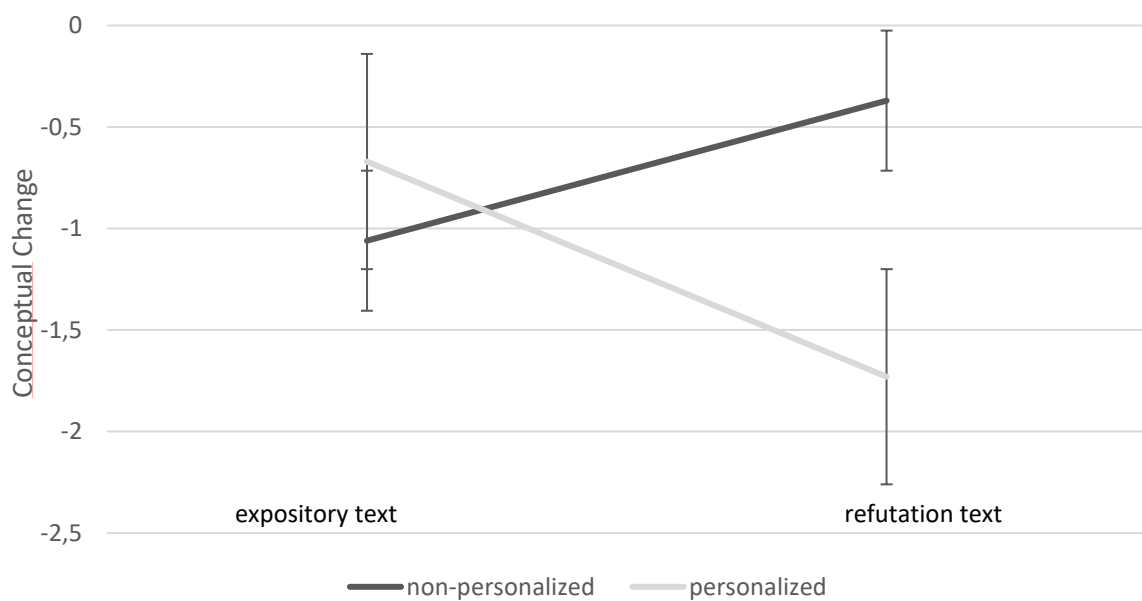
**Table 2**  
*Descriptive Values for Misconception Items*

<b>Empathizing-Systemizing (ES):</b> $\omega = .84$	Agreement rates (min. = -1, max. = 1)	Response certainty (min. = 0, max. = 4)	Misconception score (min. = -4, max. = +4)
ES1: <i>Mathematical correlations are usually easier for boys to understand than for girls, because boys tend to think in systematic contexts</i>	-.17	2.57 (0.88)	-0.63 (2.65)
ES2: <i>As boys, more likely think in systematic categories, they have better cognitive prerequisites for math than girls do.</i>	-.24	2.49 (0.93)	-0.81 (2.54)
ES3: <i>Female empathy makes it easier for girls to deal with people, while boys are on average more gifted in systematic thinking and thus in mathematics.</i>	.00	2.63 (0.85)	-0.18 (2.76)
ES4: <i>On average, girls think more empathically than boys do, boys on the other hand are more talented in systematic thinking and thus also in mathematics.</i>	.11	2.61 (0.87)	-0.17 (2.75)
<b>Girls' Compensation (GC):</b> $\omega = .76$			
GC1: <i>Girls normally have to work harder to perform equally well in mathematics as boys.</i>	-.52	2.78 (0.90)	-1.62 (2.43)
GC2: <i>Girls compensate for their on average lower aptitude in mathematics compared to boys by being more diligent.</i>	.18	2.74 (0.93)	0.47 (2.86)
GC3: <i>Mathematical content often just flies to boys, while girls have to make more effort for it on average.</i>	-.43	2.73 (0.90)	-1.41 (2.51)
GC4: <i>To achieve equally good grades in mathematics, boys have to make less effort because they are more talented than girls are.</i>	-.76	2.97 (0.87)	-1.94 (1.94)
<b>All Items:</b> $\omega = .89$			

**Figure 1***Components of the Different Manipulations (of Study Materials) Between Conditions*

**Figure 2***Result Patterns of the Personalized Refutation Text Hypothesis*

*Notes.* In this figure, we included the results of all participants who held both misconceptions in the pre-test ( $n = 96$ ). The standard error of the mean is the applied measure of error.

**Figure 3***Result Pattern of Backfire Effects for Participants Who Had Negative Conceptual Change ( $n = 42$ )*

## Appendix A

### *All Variations of Text Interventions*

		<b>Empathizing-Systemizing</b>			
		<b>Common Expository Text</b>	<b>Common Refutation Text</b>	<b>Personalized Expository Text</b>	<b>Personalized Refutation Text</b>
<b>Personalized Feedback</b>	No misconception			The evaluation of your answers shows that you have answered the questions on empathic and systematic thinking in boys and girls and their effects on mathematical talent in line with the scientific findings. In the following, you can read the current state of scientific evidence on the topic and supplement your knowledge.	
	Uncertainty			The evaluation of your answers shows that you answered the questions on empathic and systematic thinking in boys and girls and their effects on mathematical talent with uncertainty. We would therefore now like to give you the opportunity to familiarize yourself with the current state of research in this area.	
	Misconception			The evaluation of your answers shows that you have not answered the questions on empathic and systematic thinking in boys and girls and their effects on mathematical talent in line with the current scientific findings. We would therefore now like to give you the opportunity to familiarize yourself with the current state of research in this area.	

	<b>Common Expository Text</b>	<b>Common Refutation Text</b>	<b>Personalized Expository Text</b>	<b>Personalized Refutation Text</b>
<b>Refutation</b>		<p>Many people think that the male and female brains differ from birth and that boys therefore think more systematically than girls from birth. Systematic thinking is elementary for mastering mathematics. On the other hand, they assume that girls tend to think empathically, which is less helpful for mastering mathematics. Accordingly, many people assume that boys are better at mathematics from birth.</p> <p><b>However, the scientific consensus refutes this assumption.</b></p>		<p>Many people think that the male and female brains differ from birth and that boys therefore think more systematically than girls from birth. Systematic thinking is elementary for mastering mathematics. On the other hand, they assume that girls tend to think empathically, which is less helpful for mastering mathematics. Accordingly, many people assume that boys are better at mathematics from birth.</p> <p><b>However, the scientific consensus refutes this assumption.</b></p>
<b>(Alternative) Explanation</b>	<p>The same brain regions are active in different genders during the processing of math tasks (Kersey et al., 2016). This argues against structural brain differences influencing mathematical performance. Furthermore, boys and girls do not differ in mathematical ability in early childhood (Lachance &amp; Mazzocco, 2006), arguing against innate talent differences in mathematics (Girelli, 2021).</p> <p>Moreover, in a study that met the highest scientific standards, no relationship was found between systematic thinking and mathematical ability (Escovar et al., 2016). Thus, even if boys tend to be more systematic thinkers, this does not predict better mathematics performance. Finally, empathic and systematic thinking cannot be perceived as opposing or mutually exclusive abilities. This is because general intelligence provides an important foundation for both types of thinking-empathic and systematic (Gottfredson, 1998). Accordingly, individuals who think more systematically should also have higher levels of empathic thinking and vice versa.</p>			

		<b>Girls' Compensation</b>			
		<b>Common Expository Text</b>	<b>Common Refutation Text</b>	<b>Personalized Expository Text</b>	<b>Personalized Refutation Text</b>
<b>Personalized Feedback</b>	No misconception			The evaluation of your answers shows that you have answered the questions on possible compensation for mathematical talent in girls in line with the current scientific findings. In the following, you can read the current state of scientific evidence on the topic and supplement your knowledge.	
	Uncertainty			The evaluation of your answers shows that you have answered the questions on possible compensation for mathematical talent in girls with uncertainty. We would therefore now like to give you the opportunity to familiarize yourself with the current state of research in this area.	
	Mis-conception			The evaluation shows that you have not answered the questions about possible compensation for mathematical talent in girls in line with the current scientific findings. We would therefore now like to give you the opportunity to familiarize yourself with the current state of research in this area.	
<b>Refutation</b>			Many people think that girls are naturally more industrious than boys and that girls make up for their lack of innate talent in math. <b>However, the scientific evidence speaks against this assumption.</b>	Many people think that girls are naturally more industrious than boys and that girls make up for their lack of innate talent in math. <b>However, the scientific evidence speaks against this assumption.</b>	
<b>(Alternative) Explanation</b>		Girls on average invest more time in school related tasks than boys, also in mathematics (Duckworth & Seligman, 2006). However, their effort also leads to higher, not similar, mathematics grades in comparison to boys (Duckworth & Seligman, 2006). It is therefore unlikely that girls compensate for any lacking talent, since their higher investment in comparison to boys also leads to a higher performance. What further speaks against the need of compensation is that the genders, according to different findings, have similar mathematical talent: Girls and boys do not differ in terms of their mathematical abilities in early childhood (Lachance & Mazzocco, 2006). Older girls and boys also show similar performance in mathematical achievement tests (Lindberg et al., 2010). Small differences in performance can be attributed to social influences such as the role of women in the researched societies (Reilly et al., 2012). Additionally, social factors such as financial poverty or parental education have a much greater influence on performance differences in mathematics than gender (PISA, 2018). All of these findings speak against the need for compensation, as there do not appear to be any innate differences in talent to the disadvantage of girls.			

*Note.* Texts were translated from German

### List of Publications

- Dersch, A.-S.**, Platte, S. J., Watt, H., & Eitel, A. (submitted). Thinking is not Telling: Student Teachers' Implicit Math-Gender Association is Unrelated to their Explicit Math-Gender Stereotype.
- Hataru, V., Norris, C. M., **Dersch, A.-S.**, & Fauskanger, J. (submitted). Pre-service secondary mathematics teachers' views about teacher professional knowledge and potential sources of this knowledge
- \*Dersch, A.-S.**, Thacker, I., & Eitel, A. (submitted). Should We Tell Teachers That They Are (In)correct? How Personalized Refutation Text Promotes or Hampers Conceptual Change.
- \*Dersch, A.-S.**, Heyder, A., & Eitel, A., (submitted). Teachers' Awareness of Math-gender Stereotype-reinforcing Behaviors in the Classroom: A Video Experiment.
- \*Dersch, A.-S.**, Heyder, A., & Eitel, A. (2022). Exploring the Nature of Teachers' Math-Gender Stereotypes: The Math-Gender Misconception Questionnaire. *Frontiers in psychology*, *13*, 820254.  
<https://doi.org/10.3389/fpsyg.2022.820254>
- Dersch, A.-S.**, Renkl, A., & Eitel, A. (2022). Personalized refutation texts best stimulate teachers' conceptual change about multimedia learning. *Journal of Computer Assisted Learning*, *38*(4), 977-992.  
<https://doi.org/10.1111/jcal.12671>
- Eitel, A., **Dersch, A.-S.** & Renkl, A. Wissenschaftliche Annahmen und Fehlkonzepte angehender Lehrkräfte über das Lernen mit Multimedia. *Unterrichtswiss.*, *47*, 451–474 (2019). <https://doi.org/10.1007/s42010-019-00049-4>

\*Manuscripts that are part of this dissertation are marked with an asterisk.

### **Declaration of Originality**

I hereby declare that I have prepared the thesis at hand independently and without undue aid or the use of any resources other than indicated within the thesis. All parts of my thesis taken either verbally or analogously from the published or unpublished works of or based on oral communications with others are indicated as such. Regarding all aspects of my scientific enquiries as they appear in my thesis, I have upheld the tenets of good scientific practice as laid out in the "Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis" and complied with the precept of ethics, data protection and animal welfare. I declare that I have neither directly nor indirectly given monetary or any other valuable considerations to others in connection with the thesis at hand. I declare that I have not presented the thesis at hand, either in an identical or similar form, to an examination office or agency in Germany or any other country as part of any examination or degree. All materials from other sources as well as all works performed by others used or directly referenced within the thesis at hand have been indicated as such. In particular, all persons involved directly or indirectly in the development of the thesis at hand have been named. I agree with the screening of my thesis for plagiarism via offline or online detection software.

---

Anna-Sophia Dersch